# STRUCTURAL AND FUNCTIONAL STUDIES OF A NOVEL PHOSPHOTYROSINE-BINDING DOMAIN IN HAKAI, THE HYB DOMAIN, THAT TARGETS E-CADHERIN AND OTHER SRC SUBSTRATES

**MANJEET MUKHERJEE**

**NATIONAL UNIVERSITY OF SINGAPORE**

**2014**

# STRUCTURAL AND FUNCTIONAL STUDIES OF A NOVEL PHOSPHOTYROSINE-BINDING DOMAIN IN HAKAI, THE HYB DOMAIN, THAT TARGETS E-CADHERIN AND OTHER SRC SUBSTRATES

**MANJEET MUKHERJEE**
**(B.E.)**

**A THESIS SUBMITTED**

**FOR THE DEGREE OF DOCTOR OF PHILOSOPHY**


**DEPARTMENT OF BIOLOGICAL SCIENCES**

**NATIONAL UNIVERSITY OF SINGAPORE**

**2014**

# DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety.

I have duly acknowledged all the sources of information, which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

_____

Manjeet Mukherjee
1 February 2014

*I dedicate this thesis to my family*

# Acknowledgements

I would like to express my earnest thanks and sincere gratitude to my supervisor, Prof. J. Sivaraman, for his patient guidance, profound technical expertise, constant enthusiasm and unconditional support throughout my PhD. Without him, this task would have been impossible. I am truly obliged to him and hope to continue to receive his inspiring association in future.

I also express my sincere gratitude to our collaborator Prof. Graeme Guy from the Institute of Molecular and Cell Biology (IMCB), Singapore for the mammalian cell-based experiments. I thank Dr. Permeen Yusoff and Dr. Chow Soah Yee from IMCB for their contributions in the cell-based studies. I thank Dr. Fan Jing-song for his help and discussion on the NMR experiments. I thank A/P Siu Kwan Sze from NTU for his contributions in the proteomics studies. I would like to convey my sincere appreciation to Prof. K. Swaminathan for his constant encouragement, enthusiasm and valuable discussions. His positive attitude, eagerness to teach and share his knowledge, and passion for crystallography inspires me. I am thankful to Prof. Low Boon Chuan for his scientific advice and insightful discussions. I thank Prof. Ganesh Anand for the helpful discussions. I acknowledge Dr. Chiradip Chatterjee for his friendship, help and advice.

My profound thanks and gratitude goes to a number of colleagues and friends without whose support all these would not have been possible. First of all, I would like to thank Ms. Kho Say Tin and Ms. Lissa Joseph who have helped me in all the stages of my PhD by enabling the timely purchasing and procurement of the chemicals and

other resources required for my projects. I thank all members of my laboratory; Dr. Jobichen Chacko, Dr. Xingding, Dr. Xuhua, Dr. Kumar, Rajesh, Cherlyn, Veerendra, Pankaj, Nilofer, Priyanka, Mallika, Li Mo and Jeremy. I specially acknowledge Tzer Fong who taught me cloning and basic molecular biology techniques. My special thanks goes to Abhilash, Umar and Thangavelu for their friendship and companionship throughout the PhD. I am also thankful to Sarath and Digant for their friendship. I am very much thankful to all the past and present members of Lab 4 – Kuntal, Shiva, Sunil, Toan, Roopa, Anupriya, Shani, Deepthi, Pavithra, Madhuri and Divya. I also thank all my friends in the structural biology laboratories for their friendship and help; specially Vivek, Karthik, Girish, Bidhan, Sang, Kang Wei Tan, Rishi, Iman, Jack, Wang Wei, Bhaskar, Garvita and Shaveta.

I would like to remember my late grandparents who I am sure, would be very proud. I cannot thank enough my *Jethu,* my parents, my uncle and aunt, my elder brother Shrijit, my cousin Sanjeet, as well as, soon to be *Boudi*, for their encouragement, sacrifices, sincere love and constant support through the thick and thin. I dedicate this thesis to all of them. I also thank rest of my family.

I would like to convey my sincere thanks to my teachers and mentors, especially, Dr. H.G. Nagendra, Dr. N. Sathyanarayna and Dr. B. Gopal for their continuous support and encouragement over many years.

I thank NUS for providing the great opportunity to pursue my PhD here.

# Table of Contents

# Summary

Crosstalk between different types of posttranslational modifications particularly the multiple connections between phosphorylation and ubiquitination, is an emerging theme in eukaryotic biology. The premise of the present work is a step towards understanding structural and mechanistic basis of one such crosstalk, wherein a novel E3 ubiquitin ligase Hakai binds its target molecules, such as E-cadherin in a phosphotyrosine (pTyr) dependent manner leading to their ubiquitination and subsequent physiological outcomes. In this study, we focus on the pTyr-binding domain of Hakai, as it is the main functional domain that recognizes the pTyr motifs of the target proteins and act as a link for the crosstalk between the phosphorylation and ubiquitination machineries.

Chapter I provides a general introduction about two important posttranslational modification machineries, *viz.* tyrosine phosphorylation and ubiquitination, and the different elements that play central roles to mediate these posttranslational modification processes. Besides, we elaborate the novel crosstalk between phosphorylation and ubiquitination with specific examples of E3 ubiquitin ligases like c-Cbl and Hakai, which contains pTyr-binding domain, that mediate the crosstalk. Our analysis based on the sequence and knowledge gained from our previous studies and existing literature suggest the presence of a novel pTyr-binding domain in Hakai.

Chapter II describes the structural and functional characterization of the novel pTyr-binding domain of Hakai. Phosphotyrosine-binding domains, typified by the SH2 and PTB domains, are critical upstream components of signal transduction pathways. The

E3 ubiquitin ligase Hakai targets tyrosine phosphorylated E-cadherin *via* an uncharacterized domain. In this study, the crystal structure of Hakai (aa 106-206) revealed that it forms an atypical, zinc-coordinated homodimer by utilizing residues from the phosphotyrosine-binding domain of two Hakai monomers. Hakai dimerization allows the formation of a pTyr-binding pocket that recognizes specific phosphorylated tyrosines and flanking acidic amino acids of Src substrates, such as E-cadherin, cortactin and DOK1. NMR and mutational analysis identified the Hakai residues required for target binding within the binding pocket, now named the HYB domain. ZNF645 also possesses a HYB domain but demonstrates different target specificities. The HYB domain is structurally different from other phosphotyrosine-binding domains and is a potential drug target due to its novel structural features.

Chapter III describes the characterization of a C-terminal deletion mutant of HYB domain ($HYB^{\Delta C}$) that exists as a monomer in solution. The NMR structure of $HYB^{\Delta C}$ reveals that the presence of a C2H2-like zinc figure in the C-terminal domain of $HYB^{\Delta C}$ instead of the atypical zinc-coordination, results in a dramatic change in the structure of $HYB^{\Delta C}$ compared to the HYB domain, and leads to a monomeric conformation. A study on the pTyr-binding property of $HYB^{\Delta C}$ through isothermal titration calorimetry suggested that the pTyr ligand induces dimerization of $HYB^{\Delta C}$. The ligand-induced dimerization of $HYB^{\Delta C}$ is further validated through additional solution studies. These observations suggest that dimerization is absolutely necessary for the pTyr-binding property of Hakai.

The overall conclusions and the future directions are provided in Chapter IV.

# List of Tables

# List of Figures

# List of Abbreviations

| | |
|---|---|
| Å - | Angstrom ($10^{-10}$m) |
| AUC- | Analytical Ultracentrifugation |
| c-Src- | Cellular Src |
| Cbl- | Casitas B-lineage lymphoma proto-oncogene |
| DOK1- | Docking protein 1 |
| DTT- | Dithiothreitol |
| E-cadherin- | Epithelial- cadherin |
| *E.coli*- | *Escherichia Coli* |
| EMT- | Epithelial–Mesenchymal Transition |
| GST- | Glutathione S-Transferase |
| HEK293- | Human Embryonic Kidney 293 cells |
| HSQC- | Heteronuclear Single Quantum Coherence |
| HYB- | Hakai phosphotyrosine binding domain |
| IPTG- | Isopropyl β-D-1-thiogalactopyranoside |
| KDa- | Kilodalton |
| MBP- | Maltose Binding Protein |
| MDCK- | Madin-Darby Canine Kidney Epithelial Cells |
| NMR- | Nuclear magnetic resonance |
| NOESY- | Nuclear Overhauser effect spectroscopy |
| PDB- | Protein Data Bank |
| PEG- | Polyethylene Glycol |
| PKCδ- | Protein Kinase C delta |
| PKM2- | Pyruvate Kinase M Isoenzyme 2 |

| | |
|---|---|
| PTB- | Phospho-Tyrosine Binding domain |
| PTK- | Protein Tyrosine Kinase |
| PTM - | Post-Translational Modification |
| pTyr- | Phosphotyrosine |
| RAG1- | Recombination Activating Gene 1 |
| RING- | Really Interesting New Gene |
| RMSD- | Root-Mean Square Deviation |
| SAD- | Single-wavelength anomalous diffraction |
| SDS-PAGE- | Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis |
| SeMet- | Selenomethionine |
| SH2- | Src Homology 2 domain |
| TEV- | Tobacco Etch Virus |
| TKB- | Tyrosine Kinase Binding domain |
| TRIS- | Tris(Hydroxymethyl)Aminomethane |
| Trx- | Thioredoxin |
| TyrK- | Tyrosine Kinase |
| Ub- | Ubiquitin |
| v-Src- | Viral Src |
| WCL- | Whole cell lysate |
| WT- | Wild Type |
| ZnF- | Zinc Finger |
| ZNF645- | Zinc Finger Protein 645 |

**Amino Acids**

Ala (A)- Alanine

Arg (R)- Arginine

Asn (N)- Asparagine

Asp (D)- Aspartic acid

Cys (C)- Cysteine

Gln (Q)- Glutamine

Glu (E)- Glutamic acid

Gly (G)- Glycine

His (H)- Histidine

Ile (I)- Isoleucine

Leu (L)- Leucine

Lys (K)- Lysine

Met (M)- Methionine

Phe (F)- Phenylalanine

Pro (P)- Proline

Ser (S)- Serine

Thr (T)- Threonine

Trp (W)- Tryptophan

Tyr (Y)- Tyrosine

Val (V)- Valine

# Publications

1. Structure of a novel phosphotyrosine-binding domain in Hakai that targets E-cadherin, **Mukherjee, M**., Chow, S. Y., Yusoff, P., Seetharaman, J., Ng, C., Sinniah, S., Koh, X. W., Asgar, N. F., Li, D., Yim, D., Jackson, R. A., Yew, J., Qian, J., Iyu, A., Lim, Y. P., Zhou, X., Sze, S. K., Guy, G. R., and Sivaraman, J. (2012) The EMBO journal ***31, 1308-1319.***
2. Dimeric switch of Hakai-truncated monomers during substrate recognition: insights from solution studies and NMR Structure, **Mukherjee, M**., Jing-Song, F., Sarath Ramachandran, Guy, G.R., Sivaraman, J. (2014), Journal of Biological Chemistry (Under review).

# CHAPTER I

# General Introduction

## 1.1 Cell Signalling

In order to perform its normal physiological functions, cells must sense and correctly respond to the signals from their surrounding environment. Any error in sensing and responding to these signals manifests as a disease, such as cancer, neuro-degenerative disorders, auto-immunity and metabolic disorders (Anastas and Moon, 2013; Couvineau and Laburthe, 2012; Eijkelenboom and Burgering, 2013; Ma and Malynn, 2012) . Extracellular signals that a cell responds may come in various forms, including proteins, peptides, carbohydrates, lipids and inorganic ions. These are recognized by the appropriate cell-membrane receptors, and the 'message' is passed from the membrane to its target, by means of intermediaries in a signaling cascade. In the nucleus, gene expression changes occur in response to the signals received, resulting in a physiological outcome. This process is referred to as cell signalling.

In order to ensure that the correct signal is being propagated through the cell, the signaling process is both finely tuned and highly regulated. In this context, post-translational modifications (PTMs) have evolved as the eukaryotic cell's most important regulatory mechanism in cell signalling to regulate intracellular protein networks in response to external or internal cues (Deribe et al., 2010). Site-specific modification of proteins by PTMs such as phosphorylation and ubiquitination are involved in virtually all signaling pathways that orchestrate fundamental cellular processes including cell cycle progression, apoptosis, the DNA damage response, autophagy, and metabolism (Deribe et al., 2010).

## 1.2 Phosphotyrosine signalling

During the course of evolution, phosphorylation has emerged as one of the most prominent types of post-translational modification (PTM) (Hunter, 2012). Protein phosphorylation predominantly occurs on the hydroxyamino acids serine, threonine and tyrosine, and also infrequently on histidines and cysteines (Cozzone, 1988), as well as on lysines and arginines (Ciesla et al., 2011). It has been estimated that over 30% of cellular proteins are phosphorylated, representing the human phosphoproteome (Ptacek and Snyder, 2006). Although tyrosine phosphorylation accounts for less than 1% of the phosphoproteome, it plays a disproportionately major role in disease as nearly 50% of the 90 human tyrosine kinases are implicated in cancer (Del Rosario and White, 2010). In the 30 years since its discovery, tyrosine phosphorylation has been established as a fundamentally important mechanism of signal transduction and regulation in all eukaryotic cells, governing many processes, including cell proliferation, cell cycle progression, metabolic homeostasis, transcriptional activation, neural transmission, differentiation and development, and aging (Hunter, 2009).

Signal transduction *via* tyrosine phosphorylation, or what is known as the phosphotyrosine (pTyr) signalling, is mediated by an essential triad of signalling molecules: (1) the protein-tyrosine kinases (PTKs) that add a phosphate onto substrate tyrosines (conceptually these may be perceived as the ‘writers’); (2) the protein-tyrosine phosphatases (PTPs) that remove or dephosphorylate substrates (the ‘erasers’); and (3) the modular protein interaction domains that recognize the phosphorylated ligand (the ‘readers’) and hence recruit the proteins containing these

domains to specify downstream signalling events (Figure 1.1)(Lim and Pawson, 2010).

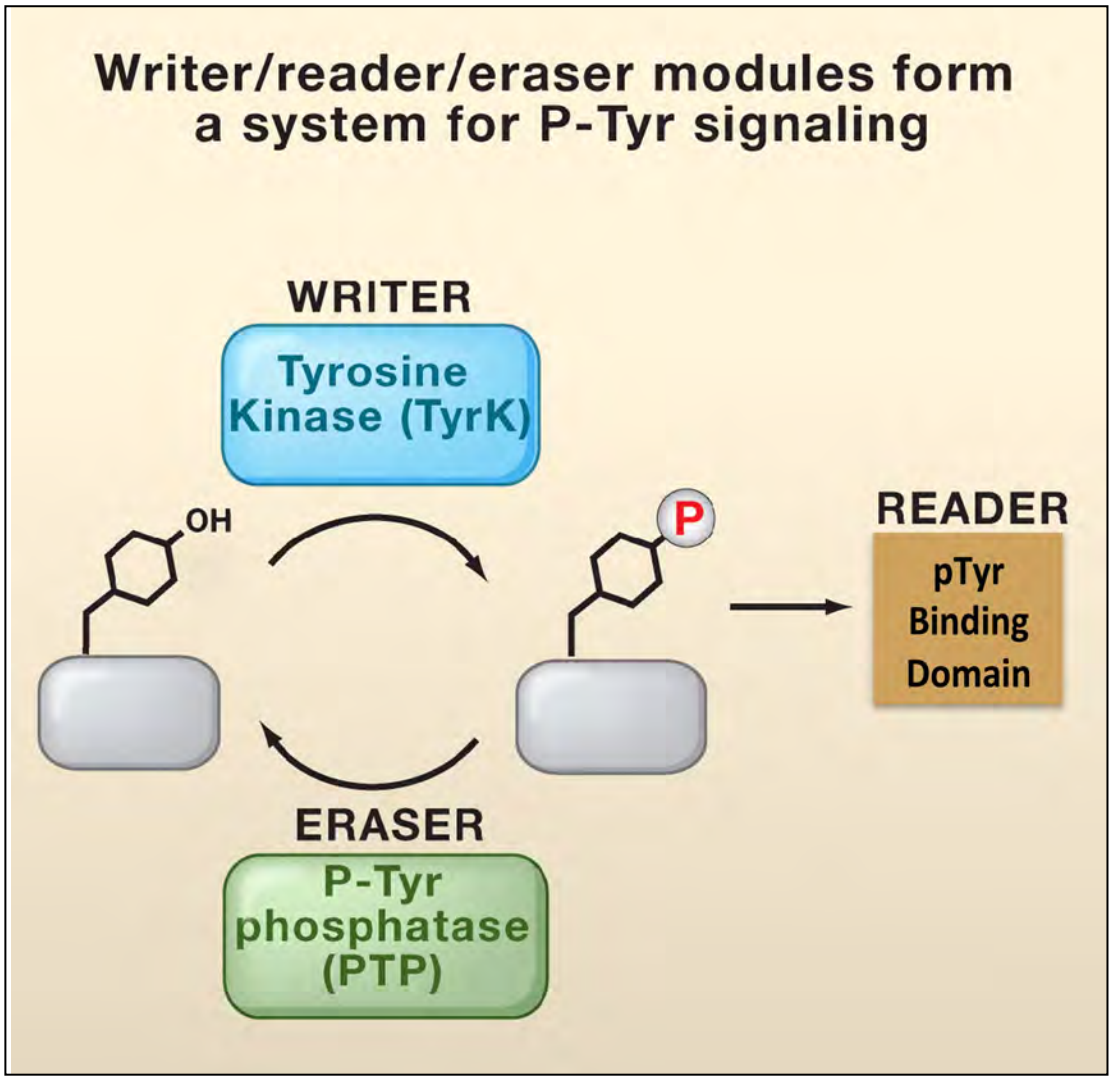



**Figure 1.1** The Writer, Reader, Eraser pTyr Toolkit. In pTyr signaling, the tyrosine kinase (TyrK), pTyr Binding , and phosphotyrosine phosphatase (PTP) domains form a highly interdependent signaling platform. This platform serves as the writer, reader, and eraser modules, respectively, for processing pTyr marks. This figure is adapted from (Lim and Pawson, 2010), with permission from Elsevier."

Protein tyrosine kinases (PTKs) catalyze the transfer of the γ phosphate of ATP to tyrosine residues on protein substrates and create binding sites for the recruitment of downstream signaling proteins (Liu and Nash, 2012). Two classes of PTKs are

present in cells: the receptor Tyrosine Kinases (RTKs) and the non-receptor Tyrosine Kinases (NRTKs). RTKs consist of an extracellular portion that binds polypeptide ligands, a trans-membrane helix, and a cytoplasmic portion that possesses tyrosine kinase catalytic activity. NRTKs, on the other hand, lack receptor-like features such as extracellular ligand-binding domain and trans-membrane-spanning region, and mostly remains localized in the cytoplasm. In addition to a tyrosine kinase domain, NRTKs possess domains that mediate protein-protein, protein-lipid, and protein-DNA interactions (Hu and Hubbard, 2005).

Protein tyrosine phosphatases (PTPs) counteract PTKs by removing phosphate moieties on target proteins. Together, PTPs and PTKs govern the timing, localization and longevity of phosphorylation events.

PTKs and PTPs commonly have somewhat limited substrate specificity and thus have limited deterministic control over downstream events (Liu and Nash, 2012). Specificity in pTyr signalling is thus largely determined by pTyr binding domains that recognize specific tyrosine phosphorylated motifs created by PTKs and act to nucleate complex formation to propagate signalling. These mainly include the Src Homology 2 (SH2) and Phospho-Tyrosine Binding (PTB) domains, which have the capability to recognize specific pTyr containing motifs (Forman-Kay and Pawson, 1999; Pawson and Nash, 2003; Yaffe, 2002). The human genome encodes roughly 90 PTKs, 107 PTPs, 54 PTB domain-containing proteins (only about one-fifth of PTB domains can bind pTyr) and 111 SH2-domain-containing proteins (Liu and Nash, 2012), which together create a complex network of regulators that constitute pTyr-signalling circuitry (Figure 1.2).

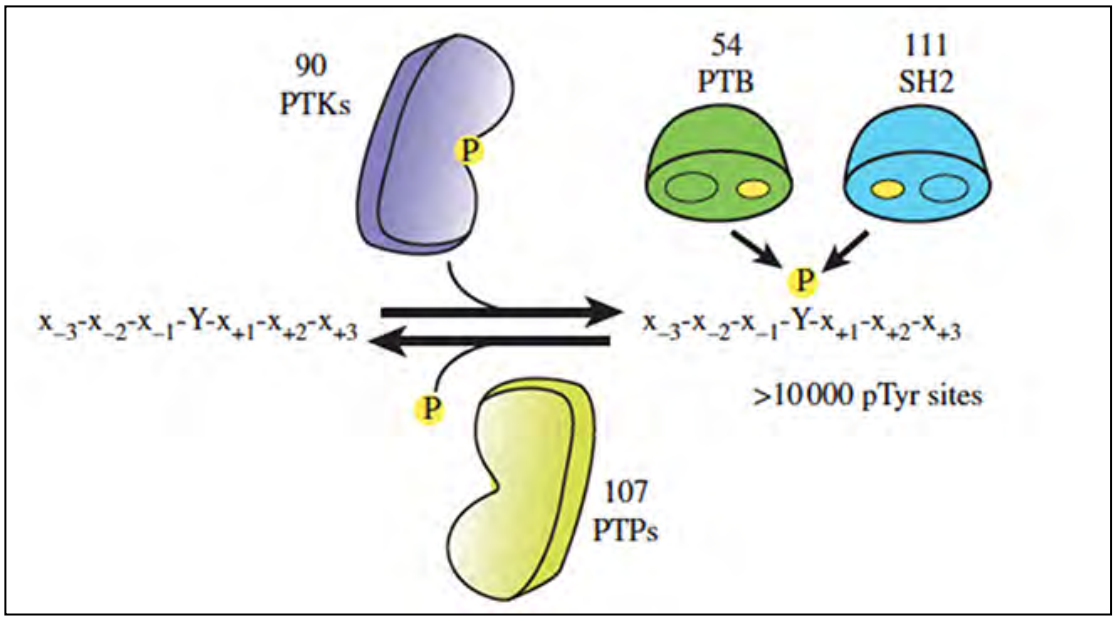



**Figure 1.2** The phosphotyrosine signalling circuitry in human. The number of writers (protein-tyrosine kinases, PTKs), erasers (protein-tyrosine phosphatases, PTP) and readers (Src homology 2, SH2; phosphotyrosine-binding domain, PTB) in humans. PTB (green) and SH2 (light blue) domain both contain a pTyr binding pocket (yellow circle), yet specificity is determined by a secondary binding site adjacent to the pTyr binding pocket that recognize sequences surrounding the pTyr. This figure is adapted from (Liu and Nash, 2012).

### 1.3 The Src kinase and its substrates

Src Kinase is a non-receptor Tyrosine Kinase and holds a remarkable historical significance for being the first Tyrosine Kinase to be discovered (Courtneidge, 2002; Martin, 2001). Src protein tyrosine kinases are 52-62 kDa proteins composed of five distinct functional domains: a unique domain, SH3, SH2, SH1 and a C-terminal regulatory region (Figure 1.3) (Boggon and Eck, 2004). The N-terminal region contains the myristylation signals that guide the Src molecule to the cell membrane while the unique domain involves in specific interaction with particular receptors and protein targets (Thomas and Brugge, 1997).

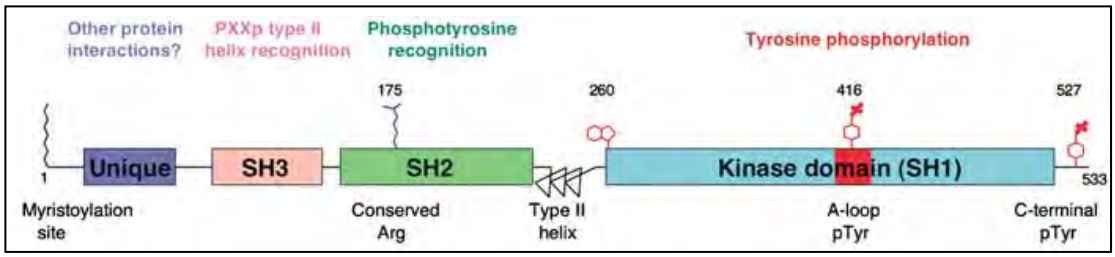



**Figure 1.3** The domain structure of Src family kinases. The Src kinase architecture consists of four domains: the unique region, which varies among family members, followed by the SH3, SH2, and tyrosine kinase domains. By convention, amino-acid residues are numbered as in chicken Src. This figure is adapted from (Boggon and Eck, 2004), with permission from Macmillan Publishers Ltd.

There are nine Src Kinases in the human genome. These proteins are all closely related to each other, and share the same regulatory mechanism. The prototype member of the Src family protein tyrosine kinases was originally identified as the transforming protein (v-Src) of the oncogenic retrovirus, Rous sarcoma virus, RSV (Brugge et al., 1977; Purchio et al., 1978). Viral v-Src is a mutated and activated version of a normal cellular protein (c-Src). V-Src always exists in opened, active conformation, whereas c-Src is flexible and normally inactive (Figure 1.4) (reviewed by (Thomas and Brugge, 1997).

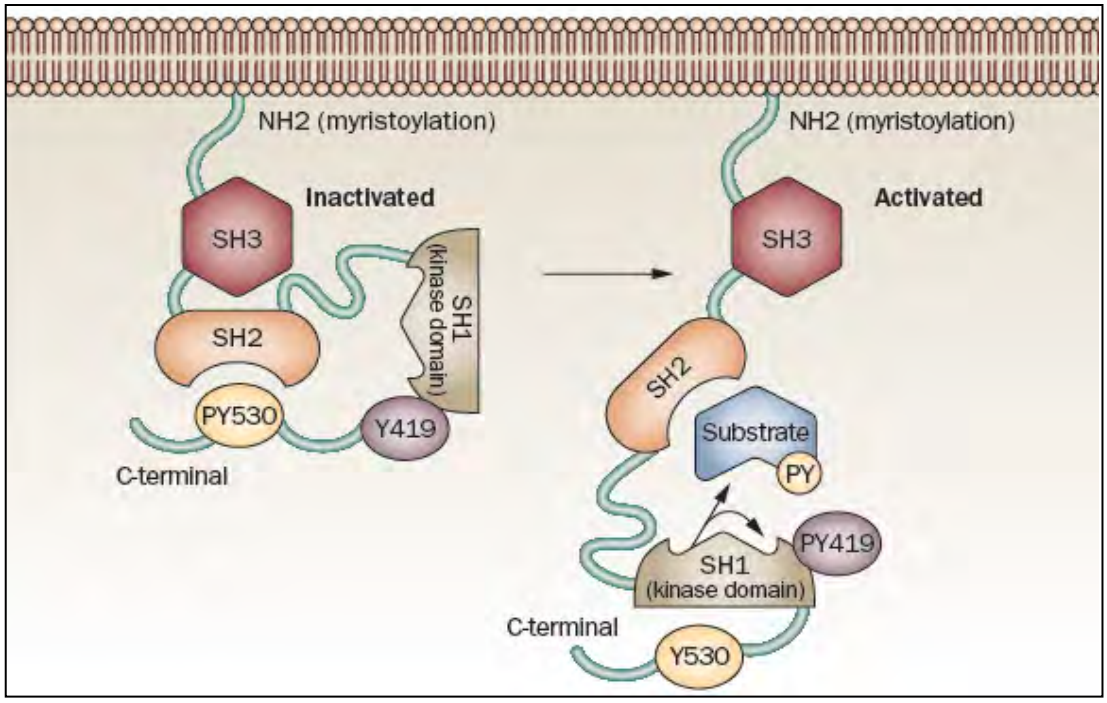



**Figure 1.4** Schematic representation of Src in the inactive (left) and the active (right) states. The left panel represents the inactive conformation of Src, in which Tyr530 interacts with the SH2 domain, positioning the SH3 domain to interact with the linker between the SH2 and catalytic domains. The right panel represents the open or active conformation. This figure is adapted from (Kim et al., 2009), with permission from Macmillan Publishers Ltd.

Src participates in the maintenance of normal cell homeostasis and in a vast range of physiological functions including cell proliferation and survival, regulation of the cytoskeleton, cell shape control, maintenance of normal intercellular contacts, cell–matrix adhesion dynamics, motility, and migration (Thomas and Brugge, 1997; Yeatman, 2004).

Src targets numerous protein substrates by creating specific pTyr sites in them to mediate the downstream signaling cascades. A recent proteomic study (Luo et al., 2008) revealed that Scr Kinases create more than 560 distinct pTyr sites on 374 protein substrates involved signaling, adhesion and cytoskeletal regulation. Activation

of Src is critically involved in carcinoma cell migration and metastasis (Sakamoto et al., 2001) (Figure 1.5). Some of the important Src substrates that are of interest in the present work include E-cadherin, Cortactin and DOK1; and are briefly discussed in the following sections.

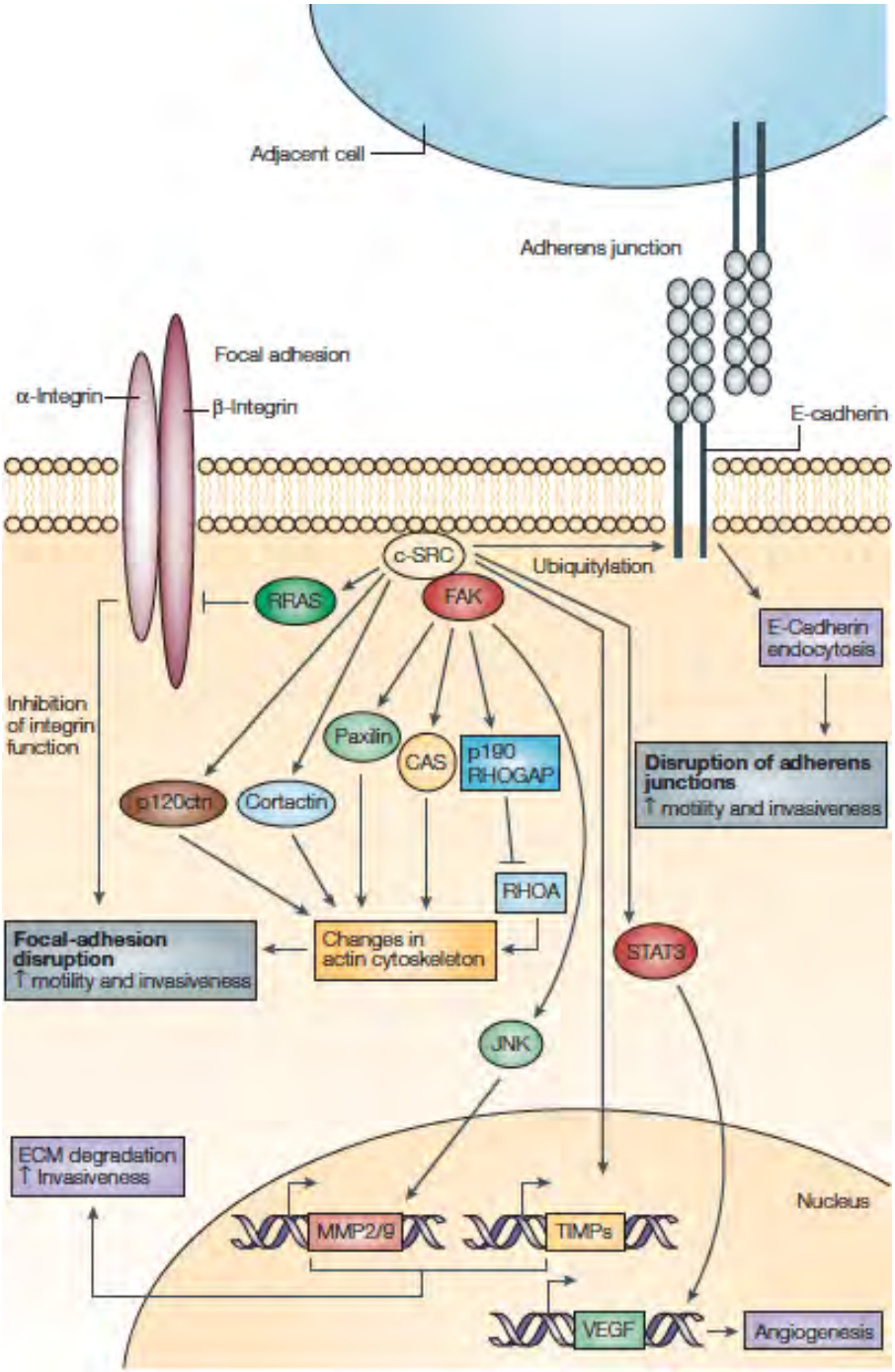



**Figure 1.5** Effects of c-SRC on tumour-cell behaviour. c-SRC exerts its effects on

tumour-cell behavior through a range of mechanisms mediated by interactions with various substrates and binding partners. A selection of these mechanisms is illustrated here. Motility and invasiveness depend on the loss of cell–cell adhesion mediated by E-cadherin. c-SRC promotes this by stimulating the ubiquitylation of E-cadherin, leading to its endocytosis. Turnover of focal adhesions is also required for motility and invasiveness, and c-SRC promotes this in various ways. Several of these mechanisms are mediated by the binding and activation of FAK, which phosphorylates substrates such as paxillin, CAS and p190 RHOGAP to bring about changes in the cytoskeleton that lead to focal-adhesion disruption. c-SRC also brings about similar changes independently of FAK through its interactions with cytoskeleton-associated proteins such as p120 catenin (p120ctn) and cortactin. Phosphorylation of RRAS by c-SRC inhibits integrin function, which also leads to focal-adhesion turnover. c-SRC activity also leads to changes in the expression of several genes that contribute to tumour progression. Activation of FAK stimulates the c-JUN amino-terminal kinase (JNK) signalling pathway, ultimately leading to increased expression of the matrix metalloproteinases MMP2 and MMP9; c-SRC also induces the expression of various tissue inhibitors of metalloproteinases (TIMPs). MMPs promote the breakdown of the extracellular matrix (ECM) that is required for tumour invasion of surrounding tissues. Signal transducer and activator of transcription 3 (STAT3) activation by c-SRC leads to increased expression of vascular endothelial growth factor (VEGF), a signalling molecule that promotes tumour angiogenesis. This figure is adapted from (Yeatman, 2004) with permission from Macmillan Publishers Ltd.

### 1.3.1 Src substrate-E-cadherin

E-cadherin is a transmembrane cell surface molecule that has a key function in epithelial cell adhesion. (Ishiyama et al., 2010; Pece and Gutkind, 2002). It consists of the extracellular cadherin domains 1–5 (EC1–5), the transmembrane region (T) and the cytoplasmic tail, which contains the juxtamembrane domain (JMD) and catenin-binding domain (CBD) (Figure 1.6). E-cadherins mediate cell adhesion by participating in the formation of adherens junctions through calcium- dependent homophilic interactions of their extracellular domains, and their cytoplasmic tail is linked to the actin cytoskeleton through p120-catenin, β-catenin and α- catenin. (Figure 1.7). β-catenin interacts with the CBD and functionally links cadherins with the actin cytoskeleton through α-catenin (Yamada et al., 2005), while p120-catenin is responsible for stabilizing cadherin-catenin complexes at the cell surface by

interacting with the JMD (Figure 1.6 and Figure 1.7) (Davis et al., 2003; Ireton et al., 2002; Lampugnani et al., 1997; Thoreson et al., 2000; Xiao et al., 2003).

The loss of E-cadherin is a major hallmark of tumor malignancy (Hanahan and Weinberg, 2000). Src plays a major role in disruption of adherens junctions by suppressing E-cadherin localization and function at these crucial contact points by inducing the tyrosine phosphorylation and endocytosis of E-cadherin (Figure 1.7) (Pece and Gutkind, 2002). Therefore, activated c-SRC promotes the release of cells from each other resulting in the conversion of epithelial cells to a more migratory mesenchymal-like phenotype, and is linked to cancer cell spread (Figure 1.8) (Avizienyte and Frame, 2005; Frame, 2004).

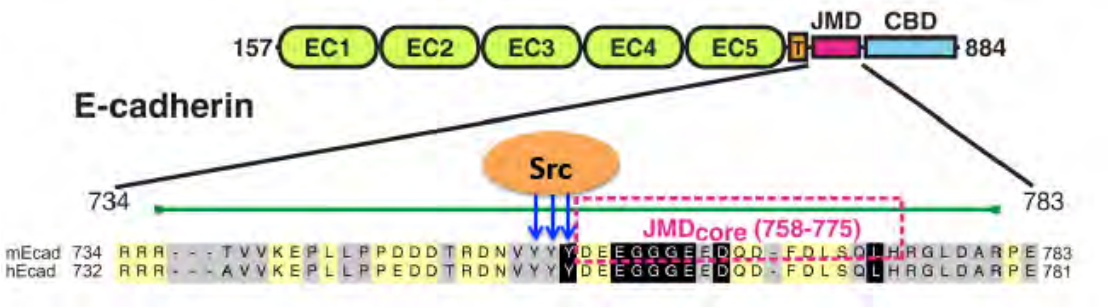



**Figure 1.6** Schematic representation of E-cadherin. E-cadherin consists of the extracellular cadherin domains 1–5 (EC1–5), the transmembrane region (T) and the cytoplasmic tail, which contains the JMD and CBD. Sequence alignment of mEcad, mouse E-cadherin and hEcad, human E-cadherin is shown. Arrows show the Src mediated tyrosine phosphorylation sites of E-cadherin. This figure is adapted from (Ishiyama et al., 2010), with permission from Elsevier.

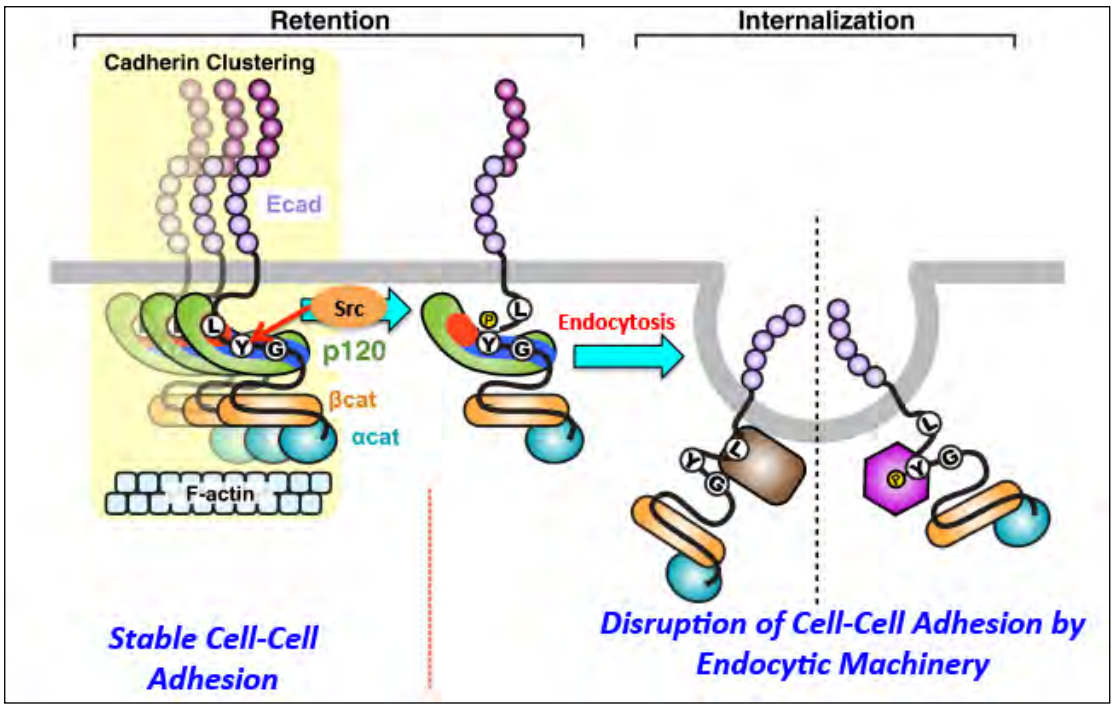



**Figure 1.7** Regulation of E-cadherin by Src. The cadherin-catenin complex consists of E-cadherin (Ecad), p120, β-catenin (βcat), and α-catenin (αcat). p120 associates with the cadherin JMD, which contains the endocytic LL motif (L), tyrosine-phosphorylation sites (Y), and triple glycine motif (G). Src phosphorylate key tyrosine residues in E-cadherin, thereby promoting their internalization by endocytosis. This figure is adapted from (Ishiyama et al., 2010), with permission from Elsevier.

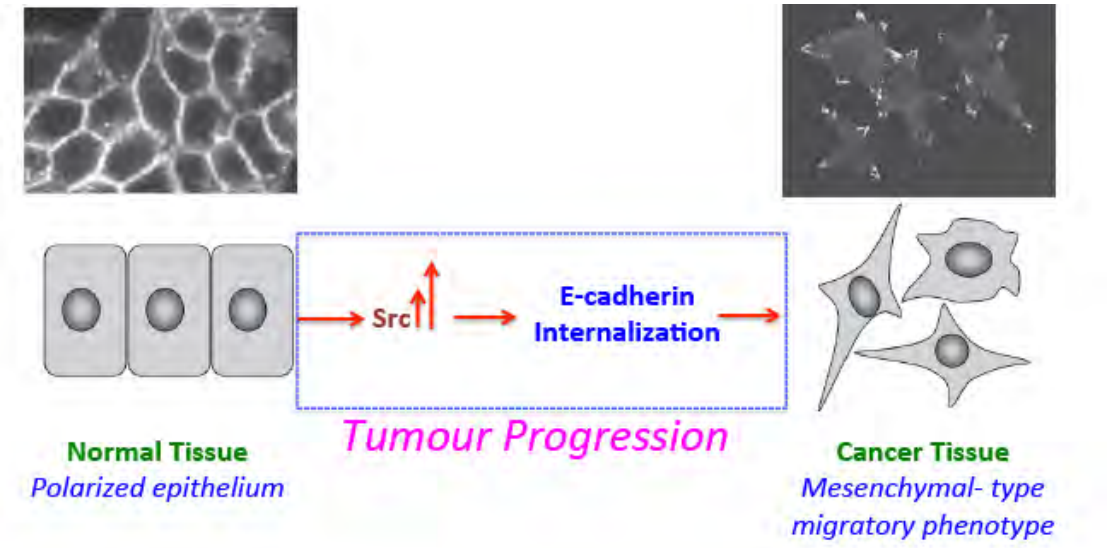



**Figure 1.8** Src induces breakdown of tissue architecture. Elevated Src activity results in suppression of cell–cell adhesion. Epithelial cells that are held together by intercellular contacts (bottom-left), and shown for cells from the KM12C colon

cancer cell series (visualized by immunofluorescence) (top-left), become dispersed as a consequence of expression of Src (bottom-right), shown for KM12C cells (top-right). E-cadherin-dependent cell-cell contacts are lost as a result of oncogenic Src, resulting in the 'adhesion switch' phenotype associated with acquisition of a more mesenchymal-like morphology as shown. This figure is adapted with permission from (Frame, 2004).

### 1.3.2 Src substrate-Cortactin

Cortactin is an actin-binding protein and a central regulator of the actin polymerization and cell migration. Cortactin was first identified as one of the major substrates for Src kinase (Wu et al., 1991) (Figure 1.9). Src affects these functions by directly phosphorylating cortactin at the leading edge where cortactin associates with and activates Arp2/3, the protein complex required to initiate actin filament polymerization and branching (Abram and Courtneidge, 2000; Rothschild et al., 2006). Cortactin is frequently overexpressed in malignant tumors and there is a direct correlation between Src mediated phosphorylation of cortactin and enhanced cell migration and metastasis (Boyer et al., 2002; Rothschild et al., 2006).

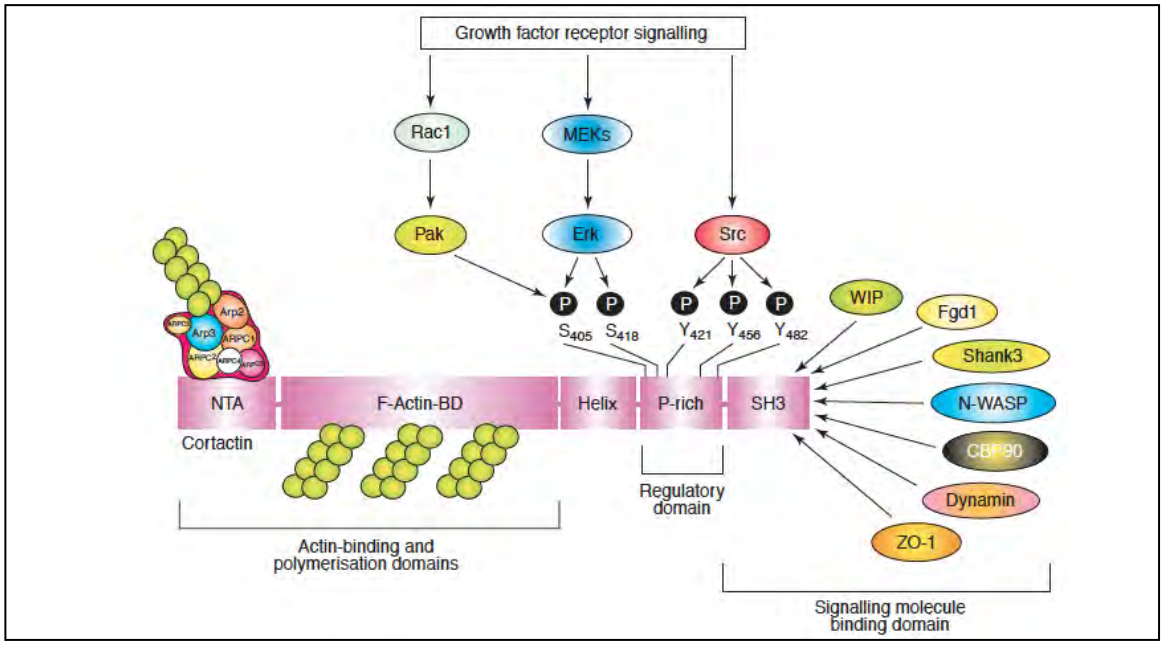



**Figure 1.9** Schematic representation of the structural domains of cortactin. The N-terminal acidic (NTA) domain can directly activate the Arp2/3 complex. The F-actin

binding domain is required for Arp2/3 activation and for the localization of cortactin at the cortical actin cytoskeleton. The proline-rich region contains serine and tyrosine phosphorylation sites and is thought to act as a molecular hinge that determines the conformation of cortactin. The SH3 domain interacts with several proteins involved in actin polymerization, cell–cell adhesion and membrane dynamics. Different signal transduction cascades regulate cortactin function *via* phosphorylation of specific serine and tyrosine residues by Pak/Erk or Src kinases, respectively. Amino-acid residues are numbered as in mouse Cortactin. This figure is adapted from (Selbach and Backert, 2005), with permission from Elsevier.

### 1.3.3 Src substrate- DOK1

DOK1 belongs to a family of adaptor proteins and serve as a scaffold protein for the recruitment of downstream signaling components. Tyrosine phosphorylation induced by Src Kinase or c-Abl are reported to be required for Dok1 functions (Lee et al., 2004; Liang et al., 2002; Rao and Mufson, 1995; Woodring et al., 2004; Zhao et al., 2001).

DOK1 is characterized by a Pleckstrin Homology domain (PH) that allows anchorage to the membrane, a PTB domain that is involved in protein-protein interactions, and a C-terminal region rich in tyrosine and serine residues (Figure 1.10) (Niu et al., 2006). Tyrosine phosphorylation by Src modulates interactions with several SH2-containing signaling molecules such as RasGAP, Nck, and the X-linked lymphoproliferative syndrome gene product SH2D1A (Niu et al., 2006).

DOK1 has emerged as a key negative regulator downstream of several receptor and non-receptor tyrosine kinase cascades. DOK1 down-regulates cell proliferation and lymphocyte signaling, inhibits mitogen-activated protein (MAP) kinase activity, and mediates activin-induced apoptosis (Kato et al., 2002; Nemorin et al., 2001; Tamir et al., 2000; Yamakawa et al., 2002). In addition to its inhibitory effects on various cellular functions, it plays positive roles in cell adhesion, cell spreading, and cell

migration (Niu et al., 2006).


**Figure 1.10** Domain architecture of DOK1. Amino-acid residues are numbered as in mouse DOK1. Tyrosine residues phosphorylated by v-Src are shown.

## 1.4 Phosphotyrosine (pTyr) binding domains in cell signalling

### 1.4.1 Major pTyr binding domains: SH2 and PTB domains

Phosphotyrosine binding domains, typified by the SH2 and PTB domains, recognize phosphotyrosine-containing motifs in a wide variety of target molecules in a sequence specific manner and enable the information flow from the cell surface to different cellular compartments to regulate the cell cycle, cell shape and movement, cell proliferation, and differentiation (Yaffe, 2002).

The SH2 domain was discovered as the first modular signalling domain that can bind to its ligands in a phosphorylation-dependent manner. (DeClue et al., 1987; Sadowski et al., 1986). At the time of discovery, 26 SH2 domain-containing proteins were predicted, but over time, more SH2 domains have been discovered and presently around 111 SH2 domains have been identified in the human genome (Liu et al., 2012; Liu and Nash, 2012). The SH2 domain is thus considered as the principal and most predominant pTyr-binding domain (Filippakopoulos et al., 2009; Forman-Kay and Pawson, 1999; Liu et al., 2012; Liu et al., 2006; Pawson and Nash, 2003; Songyang et al., 1993; Yaffe, 2002). The binding specificity of the SH2 domain is generally conferred by the sequences flanking the C-terminus of the pTyr. PTB domain was

discovered as a second conserved phosphotyrosine-binding module after the SH2 domain, identified in the adaptor protein SHC and insulin receptor substrate-1 (IRS-1) (Kavanaugh and Williams, 1994). They have since been found in various signalling molecules and currently around 54 PTB domain-containing proteins have been identified in human genome (Liu and Nash, 2012). The binding specificity of the PTB domain is generally conferred by the sequences flanking the N-terminus of the pTyr. The different mode of recognition of the phospho-peptide motif from the substrate by SH2 and PTB domains is shown in Figure 1.11. Furthermore, unlike SH2 domains which can bind only pTyr-motifs, the PTB domain also recognizes non-pTyr motifs, as evident from the observation that only about one-fifth of the 54 known PTB domains can bind pTyr (Farooq and Zhou, 2004; Forman-Kay and Pawson, 1999; Liu and Nash, 2012; Pawson and Nash, 2003; Smith et al., 2006; Yaffe, 2002).

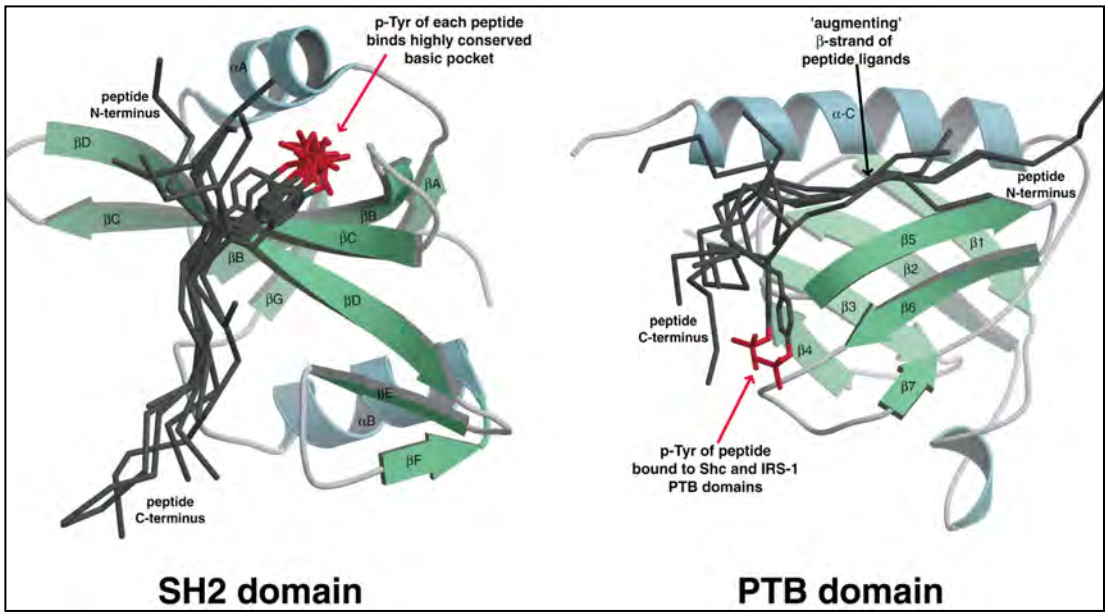



**Figure 1.11** Different modes of peptide recognition by SH2 and PTB domains. Ribbons representations are shown on the left for the Src SH2 domain [PDB code 1SPS] and on the right for IRS-1 [PDB code 1IRS]. Superimposed in each case are multiple different ligands, to show how they bind with respect to the SH2 or PTB domain fold. For SH2 domains, eight different SH2/phosphopeptide complex structures were overlaid. Peptide coordinates from each of the overlaid structures

were superimposed on a representation of the Src SH2 domain. For PTB domains, the six PTB domain/peptide complex structures were overlaid. Peptide coordinates were superimposed on a representation of the IRS-1 PTB domain. The SH2 domain structures used were Src, Lck [PDB code 1LCJ], C-terminal p85 SH2 domain [PDB code 1H9O], C-terminal PLC-γ1 SH2 domain [PDB code 2PLD], Fyn [PDB code 1AOT], Grb2 [PDB code 1TZE], SAP [PDB code 1D4W], N-terminal SH2 domain from Syp [PDB code 1AYA]. The PTB domain structures used were IRS-1 [PDB code 1IRS], Shc [PDB code 1SHC], X11 [PDB code 1X11], Numb [PDB code 1DDM], talin [PDB code 1MIZ], and radixin [PDB code 1J19]. This figure is adapted from (Schlessinger and Lemmon, 2003).

### 1.4.2 Idiosyncratic pTyr binding domains

SH2 and PTB domains are described as the two major pTyr binding domains (Forman-Kay and Pawson, 1999; Pawson and Nash, 2003; Yaffe, 2002). In addition to these, at least two more pTyr binding domains have been identified. These include the C2 domain of PKCδ (Benes et al., 2005) and the M2 isoform of Pyruvate Kinase (Christofk et al., 2008). However, both appear to be idiosyncratic properties of this particular C2 family member and Pyruvate kinase isoform respectively, as in normal situations, C2 domains are known to bind phospholipids (Benes et al., 2005) and Pyruvate Kinases catalyze the conversion of phosphoenolpyruvate to pyruvate (Christofk et al., 2008).

So far, we discussed about the various elements of pTyr mediated signalling which basically involve protein modification by the addition of small phosphate moieties on the substrate tyrosine residues and subsequent recognition of these modifications by the pTyr-binding domains which play central roles in creating a highly dynamic relay system that reads and responds to alterations in the cellular microenvironment. Another class of protein modification that plays a vital role in cell signalling involves covalent attachment of short modifier proteins such as ubiquitin and ubiquitin-like

proteins (UBLs). This process in known as Ubiquitination and is discussed in the following sections.

### 1.5 Ubiquitination

Protein ubiquitination is an important post-translational modification that involves the covalent attachment of ubiquitin and UBLs to the target proteins. Ubiquitin is a highly conserved 76-amino-acid polypeptide that is covalently attached to target proteins *via* an isopeptide bond between the carboxyl-terminal glycine of ubiquitin and the epsilon-NH2 group of a lysine in substrate proteins. Ubiquitin conjugation can affect different properties of the target (Hoege et al., 2002), including the protein's half-life, enzymatic activity, subcellular localization and protein–protein interactions (Hochstrasser, 1996; Jentsch and Pyrowolakis, 2000; Pickart, 2001; Yeh et al., 2000). It has emerged as a predominant cellular regulatory mechanism, with important roles in controlling cell division, signal transduction, embryonic development, endocytic trafficking and the immune response (Huang et al., 2004).

Ubiquitin conjugation to a substrate involves a cascade of at least three different enzymatic reactions (Figure 1.12). First, ubiquitin is activated by E1, the ubiquitin-activating enzyme, to form a high-energy thioester linkage between its C-terminal glycine residue and an active cysteine on the E1. Next, the thiol-linked ubiquitin is transiently transferred to the next enzyme in the cascade- E2, the ubiquitin-conjugating enzyme. Finally, an E3 ubiquitin ligase either transfers the activated ubiquitin molecule from the E2 to a lysine residue on the substrate or facilitates the transfer of ubiquitin from the E2 directly to the substrate (Hershko and Ciechanover, 1998). This series of events can also take the reverse course, through the action of

deubiquitinating (DUBs) enzymes, which remove ubiquitin chains from specific ubiquitin-protein conjugates (Wilkinson, 2000).

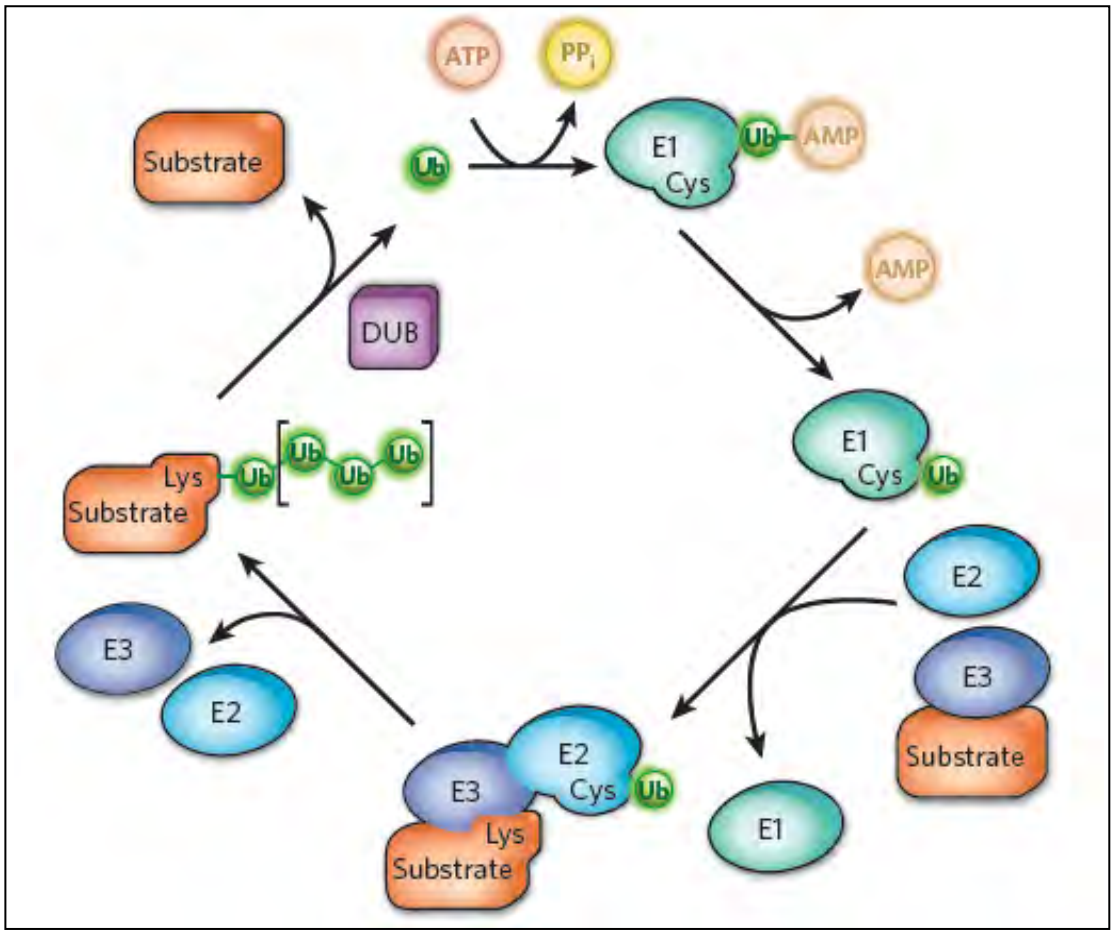



**Figure 1.12** Overview of the ubiquitin-conjugation pathway. Ubiquitin (green circles) is and attached to substrate proteins by a series of enzymes. Three types of enzyme — E1, E2 and E3 — carry out ubiquitin-modification reactions, including the assembly of polyubiquitin chains (conjugation of additional ubiquitins to a single ubiquitin to form a polyubiquitin chain on some proteins is indicated by the brackets in the image). E1s activate ubiquitin. E2s pick up the ubiquitin by transthiolation from E1 and conjugate it to substrates. E3s then ligate the ubiquitin to the substrate (and in some cases, form an intermediary thioester with ubiquitin before the final transfer to the substrate). All eukaryotes encode multiple isozymes of E2 and E3, up to several dozen E2s and many hundreds of E3s. This allows the modification of many proteins in a highly specific manner, and such modifications are often under strict temporal and spatial control. This figure is adapted from (Hochstrasser, 2009), with permission from Macmillan Publishers Ltd.

There are many forms by which the target protein could be modified by ubiquitination. The various ubiquitin modifications adopt distinct conformations and lead to different outcomes in cells. The addition of a single ubiquitin molecule to a substrate is defined as mono-ubiquitination (Figure 1.13) (Hicke and Dunn, 2003). Alternatively, several lysine residues in the substrate can be tagged with single ubiquitin molecules, giving rise to multiple mono-ubiquitination, also referred to as multi-ubiquitination (Figure 1.13) (Haglund et al., 2003). Moreover, ubiquitin contains seven lysine residues, which can also be targeted by another ubiquitin in an iterative process, known as poly-ubiquitination, that leads to the formation of a ubiquitin chain attached to a single lysine of a protein substrate (Figure 1.13) (Hofmann and Pickart, 2001). It is now clear that different types of ubiquitin conjugates are involved in the regulation of different cellular processes.

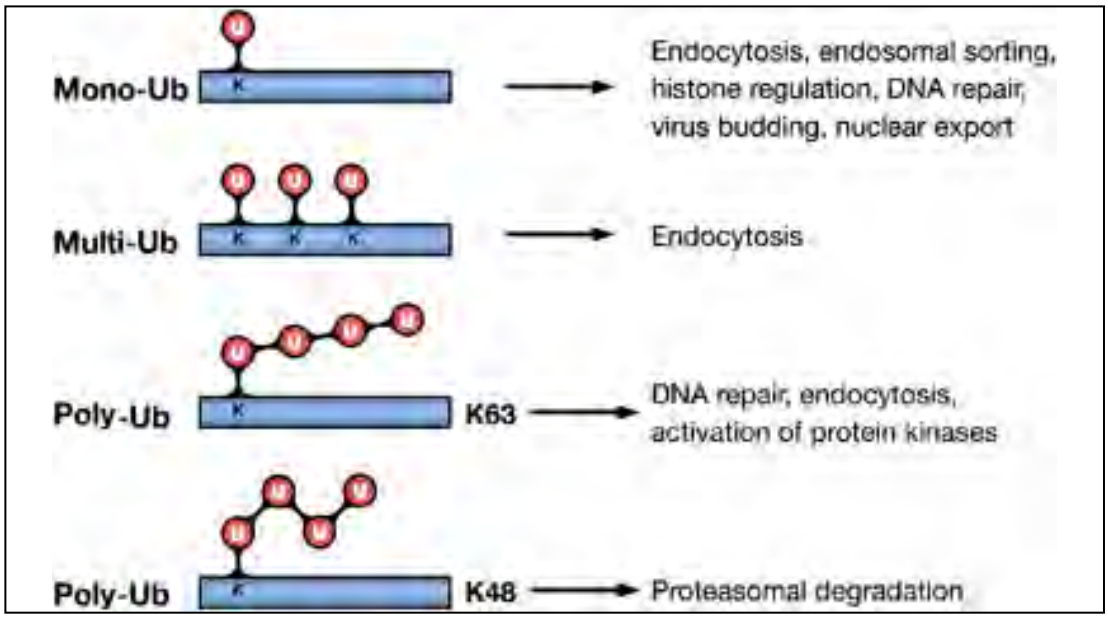



**Figure 1.13** Ub modifications and their cellular functions. Attachment of a single Ub molecule to a single Lys (K) residue leads to protein mono-ubiquitylation (mono-Ub). Multiple mono-ubiquitylation (multi-Ub) results from the addition of several single Ub molecules to different Lys residues. These modifications are implicated in various nonproteolytic cellular functions, including endocytosis, endosomal sorting and DNA repair. Poly-ubiquitylation (poly-Ub) results from the attachment of a chain of Ub

molecules to one or more Lys residues. Ub chains formed via Lys48 (K48) of Ub target-modified proteins for proteasomal degradation. Chains linked via Lys63 (K63) are implicated in DNA repair and activation of protein kinases. This figure is adapted from (Haglund and Dikic, 2005), with permission from Macmillan Publishers Ltd.

### 1.5.1 E1 and E2 enzymes

The human ubiquitin pathway involves a hierarchical network that is constituted by two E1 enzymes, 37 E2 enzymes, over 600 E3 ligases, and thousands of targets with diverse biological functions (Komander, 2009; Schulman, 2011). UBA1 and UBA6 are the two E1 enzymes present in human (Pelzer et al., 2007; Schulman and Harper, 2009). The E1 enzymes consist of three independently- folded globular domains, each specifying distinct E1 activities: (1) an adenylation domain binds the ubiquitin and MgATP and contains the adenylation reaction active site; (2) a catalytic Cys domain that harbors the active site Cys that is involved in the E1~Ubiquitin thioester linkage and (3) a carboxyl-terminal ubiquitin-fold domain (UFD) that resembles ubiquitin that binds E2 (Schulman, 2011). E1 enzymes use ATP to generate a thioester bond between the Cys at their active site and the carboxyl terminus of ubiquitin and then transfers this activated ubiquitin to the Cys residue in the active site of an E2 enzyme (Ye and Rape, 2009). E2 enzymes act *via* selective protein–protein interactions with the E1 and E3 enzymes and connect activation to covalent modification. E2s contain a highly conserved ubiquitin-conjugating (UBC) domain, which accommodate the ATP-activated ubiquitin from the E1 enzyme *via* a covalently linked thioester onto its active-site residue. After being charged with ubiquitin, E2s engage E3s to catalyse substrate ubiquitylation. A single E2 can interact with several different E3s (Komander, 2009; Ye and Rape, 2009). Although the analysis of E2–E3 interactions *in vitro* is straightforward, determining the physiological E2–E3 pairs, especially in human cells, is more difficult (Ye and Rape, 2009).

### 1.5.2 E3 ubiquitin ligases

E3 ligases confer specificity to ubiquitination by recognizing target substrates and mediating transfer of ubiquitin from an E2 ubiquitin conjugating enzyme to substrate (Deshaies and Joazeiro, 2009). There are three major types of E3 ligases in eukaryotes. These include a RING, U-box and HECT (Ardley and Robinson, 2005; Gao and Karin, 2005).

#### 1.5.2.1 RING E3 ligases

RING (really interesting new gene) E3 Ligases are the most prominent class of ubiquitin-ligases with over 600 different RING ligases encoded by human genome (Deshaies and Joazeiro, 2009). RING domain has several cysteines and a histidine in its core structure comprising of 40 to 60 amino acid residues which bind two zinc atoms (Borden and Freemont, 1996) (Figure 1.14). RING E3 ligases act as adaptor which bring an E2 bound with ubiquitin (E2~Ub) and a substrate into sufficiently close proximity and promote direct transfer of ubiquitin from E2~Ub to substrate (Deshaies and Joazeiro, 2009) (Figure 1.17). Members of the RING finger ubiquitin ligase family can function as monomers, dimers or multi-subunit complexes. Dimerization generally occurs through the RING finger domain or surrounding regions and can result in homodimers [e.g. cIAP, RNF4, SIAH, and TRAF2] (Liew et al., 2010; Mace et al., 2008; Park et al., 1999; Polekhina et al., 2002) or heterodimers [e.g. (MDM2 and MDMX), (BRCA1 and BARD1), (RING1b and BMI1)] (Metzger et al., 2012). For heterodimers, one RING domain (MDMX, BARD1, BMI1) often lacks ligase activity and might conform and/or stabilize the active E2-binding RING domain. Multi-subunit RING domains are exemplified by the cullin RING ligase

(CRL) superfamily (Petroski and Deshaies, 2005) and anaphase-promoting complex/cyclosome (APC/C) (Metzger et al., 2012).

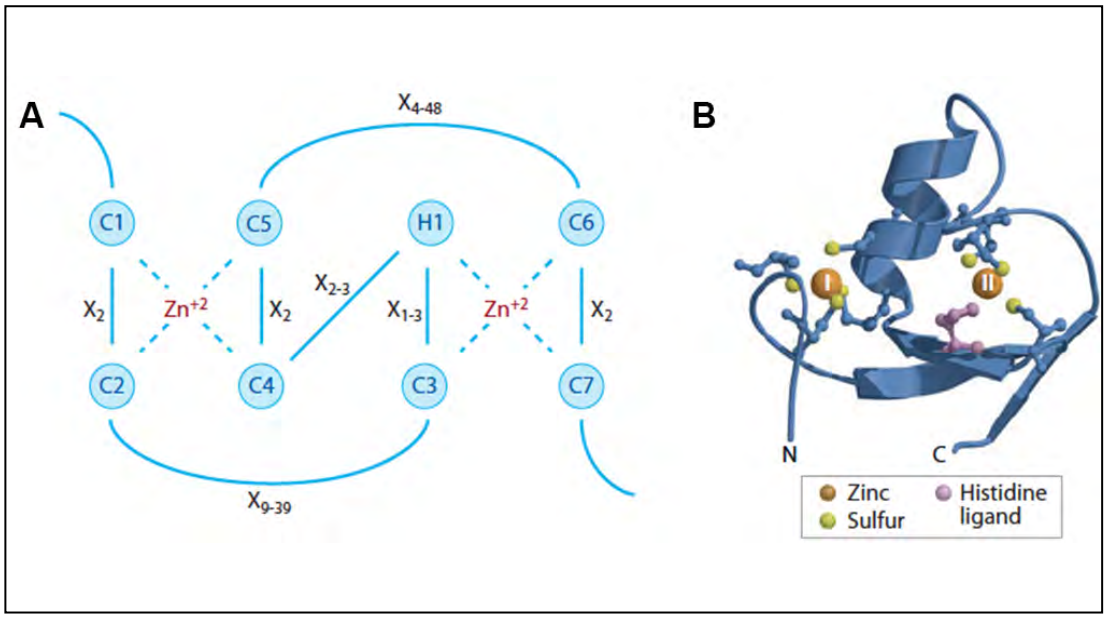



**Figure 1.14** The RING finger domain. (**A**) Primary sequence organization of the RING-HC domain. The first cysteine that coordinates zinc is labeled as C1, and so on. H1 denotes the histidine ligand. Xn refers to the number of amino acid residues in the spacer regions between the zinc ligands. (**B**) Ribbon diagram of the three-dimensional crystal structure of the RING domain from c-Cbl. The zinc atoms in sites I and II are numbered. The termini are as marked. This figure is adapted with permission from (Deshaies and Joazeiro, 2009); permission conveyed through Copyright Clearance Center, Inc.

### 1.5.2.2 U-box E3 ligases

U-box domains are structurally related to the RING and function in mediating ubiquitination (Figure 1.15 and Figure 1.17). In the U-box domain, the zinc-binding sites are replaced by conserved charged and polar residues that engage in hydrogen-bonding networks and that are required for maintaining structure and activity (Aravind and Koonin, 2000) (Figure 1.15). The human genome codes for around 7 different U-box ubiquitin ligases.

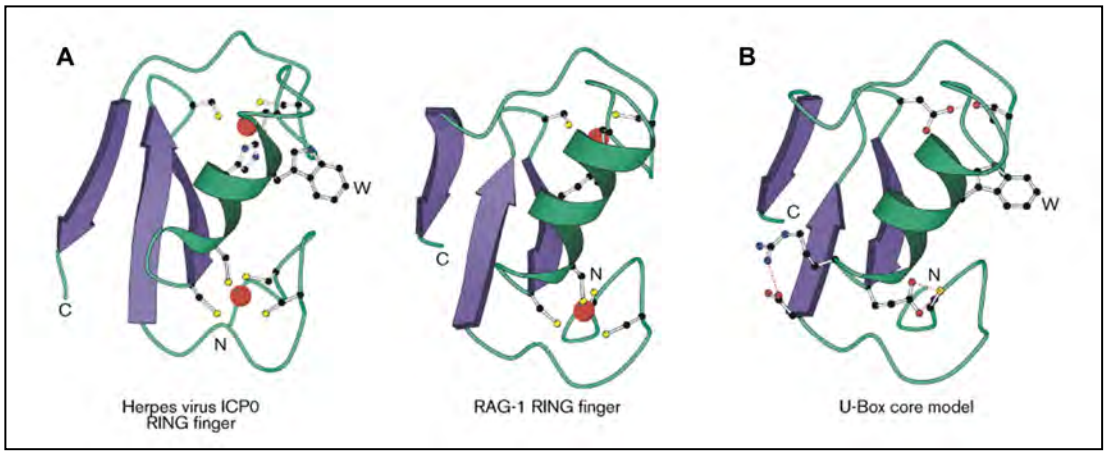



**Figure 1.15** The structural scaffold of the RING and U-box domain. (A) Known structures of the indicated protein domains. (B) Model of the U-box structure constructed using PROMODII from the structures in (A,B) and the U-box sequence from UFD2. W, Trp residue conserved in RING fingers; N, amino terminus; C, carboxyl terminus. This figure is adapted from (Aravind and Koonin, 2000), with permission from Elsevier.

### 1.5.2.3 HECT E3 ligases

HECT domains are identified on the basis of their similarity to the founding member of the family, E6AP. HECT (homologous to E6-associated protein C-terminus) E3s have a direct role in catalysis during ubiquitination. In contrast to RING domains, which can occur at any position within a given protein, all known HECT domains are found at the carboxy-terminal end of their respective proteins (Wenzel and Klevit, 2012). The HECT domain has a bilobal structure: the lobe at the amino-terminal end of the domain (the N-lobe) serves as the E2-binding domain, and the lobe at the carboxyl terminus (the C-lobe) contains the catalytic cysteine (Huang et al., 1999)(Figure 1.16). The catalytic cysteine forms thioester with ubiquitin before it is transferred from E2~Ub to the substrate. This is in sharp contrast to the RING and U-box type E3 ligases, which transfers the ubiquitin directly from E2~Ub to the substrate (Figure 1.17). Database analyses indicate that the human genome encodes 28 different HECT proteins (Scheffner and Staub, 2007).

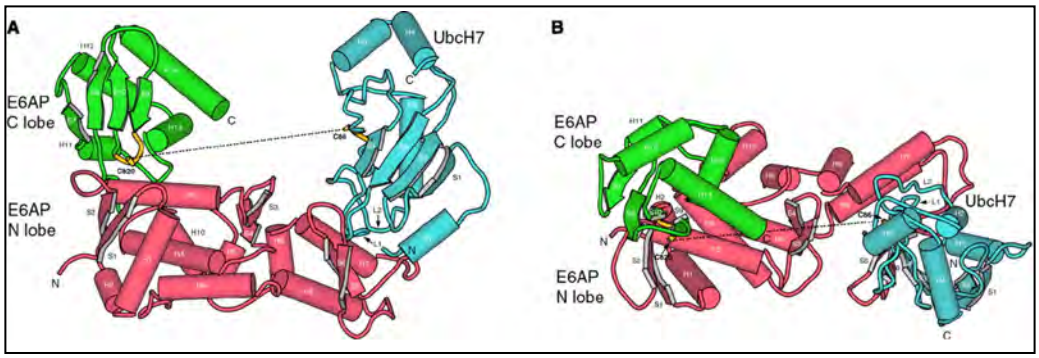



**Figure 1.16** The E6AP HECT domain–UbcH7 (UbcH7 is an E2 enzyme) complex forms a U-shaped structure. (**A** and **B**) Orthogonal views of the overall structure of the complex. The E6AP HECT domain N lobe (consisting of 12 α helices and six β strands), C lobe (six α helices and four β strands), and the E2 enzyme, UbcH7 (four α helices and four β strands) are colored in green, red, and cyan, respectively. The two active-site loops that contain the catalytic cysteine are colored yellow. The dotted line indicates the open line of sight between the active-site cysteines of E6AP and UbcH7. This figure is adapted from (Huang et al., 1999), with permission from American Association for the Advancement of Science (AAAS).

The mode of action of the different E3 ligases showing the mechanism of ubiquitin transfer from the E2~Ub to the substrate is indicated schematically in Figure 1.17.

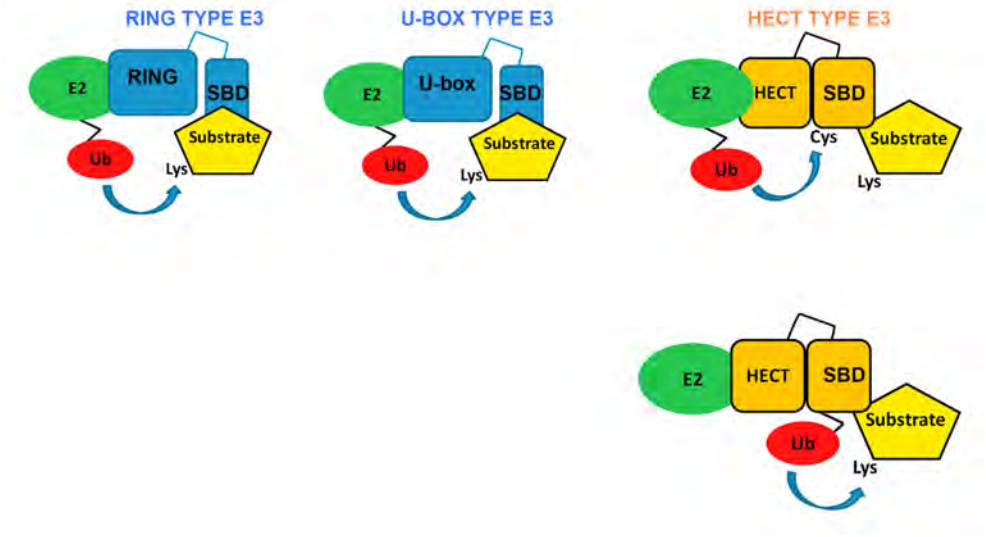



**Figure 1.17** Summary of classification of E3 ligases. RING type and U-box type E3 ligases transfer the ubiquitin directly from E2~Ub to substrate, whereas HECT type E3s form thiol-ester intermediates with ubiquitin and then transfers it to the substrate. In all the three cases, the E3 ligases engage with their substrates through their respective Substrate Binding Domain, denoted in the figure as SBD, and mediate the Ubiquitin transfer from the E2 enzyme to the substrate.

To this point, we discussed about two important PTMs- Tyrosine Phosphorylation and Ubiquitination- that play pivotal role in the regulation of cell signaling in eukaryotic cells. In recent findings, compelling amount of data indicates the involvement of a sequential series of PTMs like phosphorylation and ubiquitination in tandem in transmitting and regulating signaling (Deribe et al., 2010). In the following sections, we briefly discuss the emerging insights in the similarities and crosstalk between phosphorylation and ubiquitination.

### 1.6 Similarities between ubiquitination and phosphorylation

There are many similarities between phosphorylation and ubiquitination (Figure 1.18). First, both are reversible forms of covalent modification: just as phosphorylation can be reversed by protein phosphatases, ubiquitination can be reversed by members of a large family of enzymes known as deubiquitinating (DUBs) enzymes. Second, both occur rapidly, often within minutes or seconds following stimulation. Third, both are highly specific and controlled by a very diverse set of enzymes. The human genome project reveals the presence of >30 different E2s, >500 different E3s and >80 different Dubs (Komander, 2009). Such diversity rivals that of protein kinases (of which there are a total of 518 in humans) and phosphatases (a total of 120). Fourth, both ubiquitination and phosphorylation can be detected by specialized domains that bind to ubiquitin or phospho-amino acids, respectively. In an analogous manner as the specific sequences containing pTyr can be recognized by SH2 and PTB domains, ubiquitin and polyubiquitin chains can be recognized by several different classes of ubiquitin-binding domains present in a wide variety of

proteins (Ghosh and Karin, 2002). These domains include UIM (Ub-interacting motifs), UBA (Ub-association), CUE (Cue1-homologous), UEV (Ub E2 variant), PAZ (poly ubiquitin associated zinc finger) and NZF (novel zinc finger).

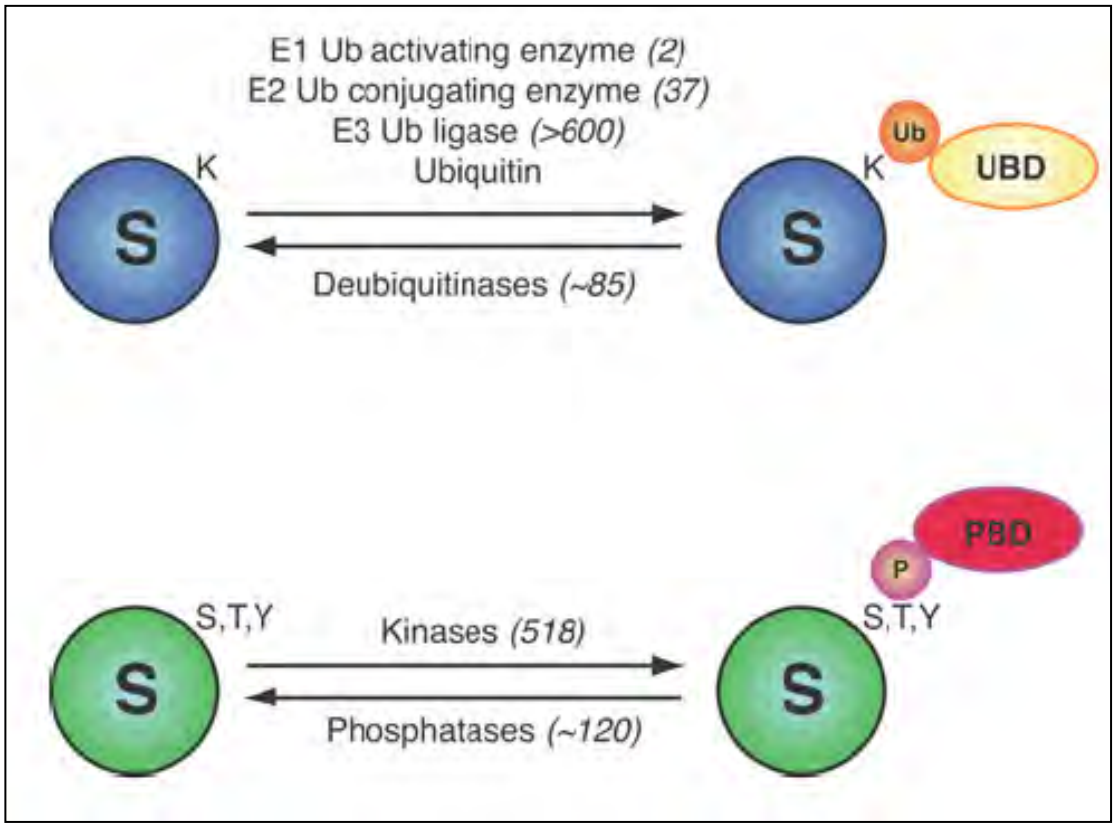



**Figure 1.18** Comparison of ubiquitination and phosphorylation Systems. Ubiquitination and phosphorylation are reversible processes. The enzymes facilitating and removing the modifications are indicated, and numbers in parentheses indicate the number of human genes encoding the respective proteins. As for protein kinases and phosphatases, E2 conjugating enzymes and deubiquitinases also comprise several pseudogenes that lack catalytic residues and are hence inactive. The roles of inactive deubiquitinases are currently unknown, whereas the inactive E2 enzymes Uev1a and MMS2 have important roles in the assembly of $Lys^{63}$-linked ubiquitin chains. K, lysine residue; P, phosphorylation; PBD, phosphate-binding domain; S, substrate; S/T/Y, serine/threonine/tyrosine residue; Ub, ubiquitin; UBD, ubiquitin-binding domain. This figure is adapted with permission from (Komander, 2009). © the Biochemical Society.

### 1.7 Crosstalk between ubiquitination and phosphorylation

Crosstalk between ubiquitination and phosphorylation occurs at several levels (Figure

1.19). Phosphorylation can promote or inhibit ubiquitination, which in turn can lead to proteasomal degradation or processing or regulate intracellular trafficking of membrane proteins (Hunter, 2007). Phosphorylation can regulate ubiquitination of a protein in three main ways. First, phosphorylation positively or negatively regulates the activity of the E3 ligase responsible for Ub transfer. Second, phosphorylation promotes recognition by an E3 ligase by creating a phosphodegron. Third, phosphorylation can influence ubiquitination by regulating substrate/ligase interaction at the level of subcellular compartmentalization.

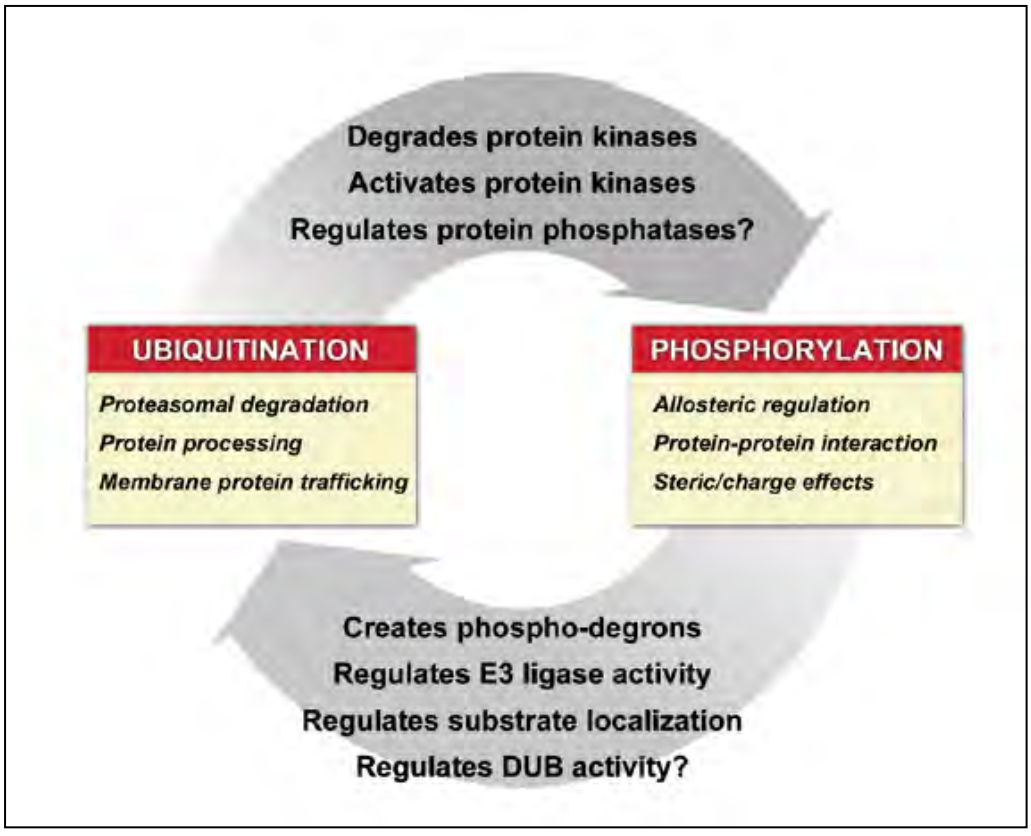



**Figure 1.19** Crosstalk between Phosphorylation and Ubiquitination. Crosstalk between phosphorylation and ubiquitination occurs at several levels (Figure 1). Phosphorylation can promote or inhibit ubiquitination, which in turn can lead to proteasomal degradation (polyUb, e.g., cyclin E) or processing (polyUb, e.g., Ci or NF-**k**B2/p100) or regulate intracellular trafficking of membrane proteins (monoUb, e.g., Ste2 and EGFR). Phosphorylation can regulate ubiquitination of a protein in three main ways. First, phosphorylation positively or negatively regulates the activity of the E3 ligase responsible for Ub transfer. Second, phosphorylation promotes

recognition by an E3 ligase by creating a phosphodegron. Third, phosphorylation can influence ubiquitination by regulating substrate/ligase interaction at the level of subcellular compartmentalization. This figure is adapted from (Hunter, 2007), with permission from Elsevier.

### 1.7.1 Example of crosstalk: pTyr dependent ubiquitination by c-Cbl

A classical example of the crosstalk between phosphorylation and ubiquitination is the phosphotyrosine dependent ubiquitination and degradation of receptor and non-receptor tyrosine kinases by the RING E3 ligase c-Cbl (Figure 1.20).

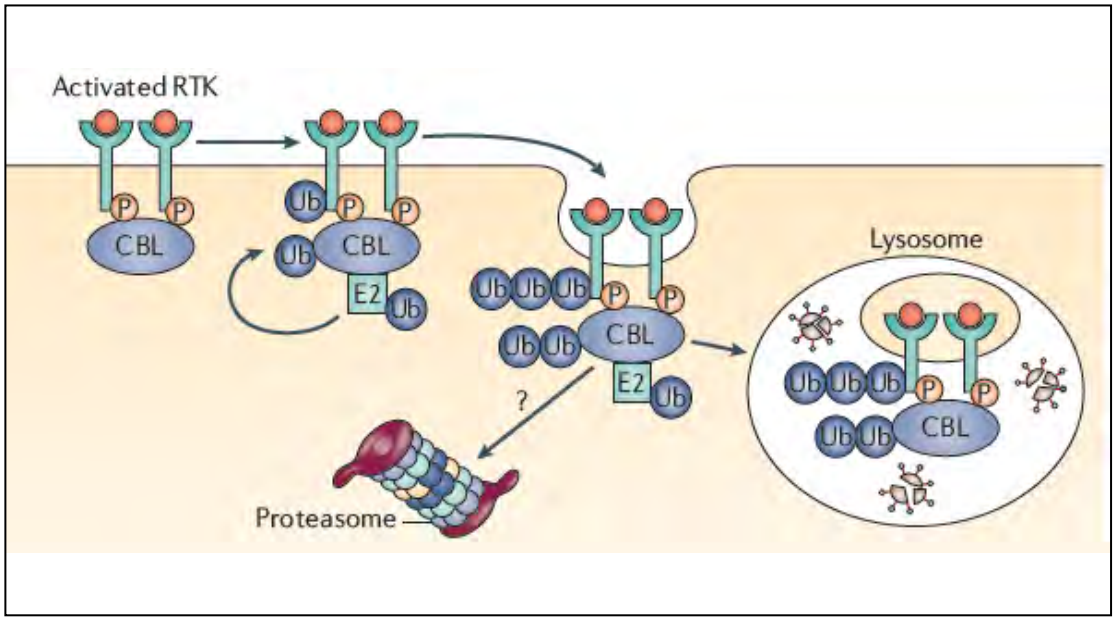



**Figure 1.20** CBL proteins mediate ubiquitylation (Ub) and the downregulation of receptor tyrosine kinases (RTKs). Whether CBL proteins associated with activated complexes are degraded in lysosomes or proteasomes is unresolved. Here, P stands for phosphorylation. This figure is adapted from (Lipkowitz and Weissman, 2011), with permission from Macmillan Publishers Ltd.

Recently, our lab has shown that c-Cbl interacts with a diverse array of RTKs and NRTKs in a phosphotyrosine dependent manner through a N-terminal pTyr- binding domain lying adjacent to the RING domain (Ng et al., 2008). Although this pTyr-binding domain appeared to have several hallmarks of a PTB consensus motif, the

crystal structure demonstrated that it structurally resembled an SH2 domain but required a flanking four-helix bundle (4H) and an EF-hand subdomain to accomplish binding (Meng et al., 1999; Ng et al., 2008). Together, these three sub-domains make up the tyrosine kinase-binding (TKB) domain, which is unique to Cbl proteins and often referred to as a specialized or 'embedded' SH2 domain (Liu et al., 2006). (Figure 1.21)

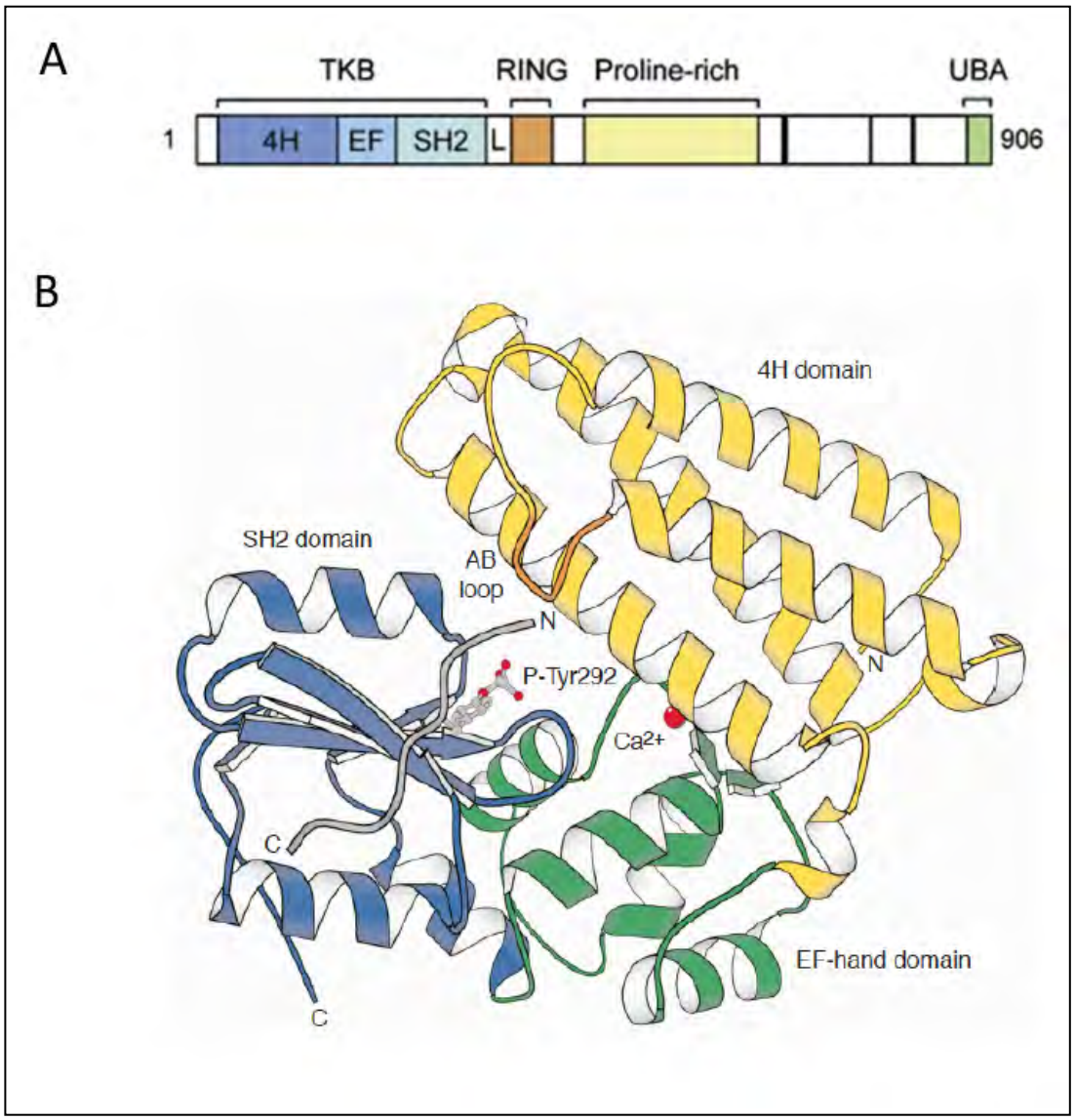



**Figure 1.21** c-Cbl and pTyr recognition. (A) Domain architecture of c-cbl [adapted from (Ng et al., 2008)] (B) Structure of the tyrosine kinase-binding (TKB) domain of Cbl bound to a ZAP-70 phosphopeptide. The Cbl TKB domain comprises three

structural domains: a four-helix bundle (yellow), an EF-hand Ca21-binding domain (green) and a divergent SH2 domain (blue). [Adapted from (Lupher et al., 1999), with permission from Elsevier.]

### 1.7.2 Unique pTyr recognition mechanism of c-Cbl TKB

The c-Cbl TKB domain binds its target proteins following phosphorylation of a central tyrosine residue in the consensus motif. Binding to the TKB domain is required for the subsequent conjugation of ubiquitin through the RING domain during the ubiquitination process (Schmidt and Dikic, 2005). The consensus binding sequence targeted by the TKB domain was originally identified by (Lupher et al., 1997) as D(N/D)XpY, but was later experimentally refined as (N/D)XpY(S/T)XXP. This motif is common to several RTKs, such as EGFR, colony-stimulating factor 1 receptor, ZAP-70, Src and Syk, as well as members of the Sprouty (Spry) family of Ras/ mitogen-activated protein kinase inhibitors (Figure 1.22). In our lab, we identified two additional consensus motifs that are targeted by the c-Cbl-TKB domain (Ng et al., 2008). The first, RA(V/I)XNQpY(S/T), is a derivation of the original sequence, with conserved residues extending back to the arginine at the pY-6 position (Figure 1.22). This sequence is conserved amongst a family of adaptor proteins based on the APS protein (Hu and Hubbard, 2005). The second consensus sequence was identified as a DpYR motif in a study investigating the binding of c-Cbl to the c-Met receptor (Met). This motif bears no resemblance to the previously characterized PTK-binding motif (Peschard et al., 2004) and is conserved among the Met family members, Ron and Sea (Penengo et al., 2003), as well as in plexins - receptors for semaphorins that promote cell repulsion (Tamagnone et al., 1999)( Figure 1.22).

Although the 111 currently documented SH2 domains use diverse strategies to

aggregate signalling complexes, there is no precedent for a given SH2 or PTB domain to associate with such apparently unrelated consensus binding motifs as seen with the TKB domain. A recent study by our lab (Ng et al., 2008) revealed the hitherto undiscovered common binding mechanism that effectively explains the basis of the unique ability of c-Cbl TKB to recognize diverse substrate motifs. By using a combination of structural and biochemical approaches, we demonstrated that TKB binding requires the formation of a unique intrapeptidyl hydrogen bond (H-bond) between the pTyr and the conserved asparagine (Asn) at the pY-2 position or an arginine (Arg) at the pY-1 position when the Asn is absent (Figure 1.23A). Furthermore, Met binds to the TKB domain in the reverse direction, from N to C terminus with the formation of an intrapeptidyl H-bond between the pTyr and the Arg at the pY-1 position (Figure 1.23B). This was the first description of an SH2 domain having the capacity to bind to proteins in two orientations. In order to gain further structural insights into the reverse binding mechanism of c-Cbl-TKB, our lab used a protein engineering approach wherein the Met peptide with the DpYR motif was changed to RpYD (MetRD) (Sun et al., 2011). Here, we demonstrated that the MetRD peptide binds with TKB with a similar binding affinity as the native Met peptide, but with a complete reversal in orientation. In this study, we established that the binding orientation is dictated by the phosphorylated tyrosine and an adjacent arginine which forms the intra-peptide hydrogen bonds and aligns the pTyr motif unidirectionally with complementary charges in the phosphotyrosine binding pocket of c-Cbl (Sun et al., 2011). Moreover, our lab has further contributed in understanding the pTyr binding mechanism of c-Cbl TKB by demonstrating that additional serine/threonine phosphorylation reduces binding affinity but preserves interface topography of substrate proteins to the c-Cbl TKB domain (Sun et al., 2010).

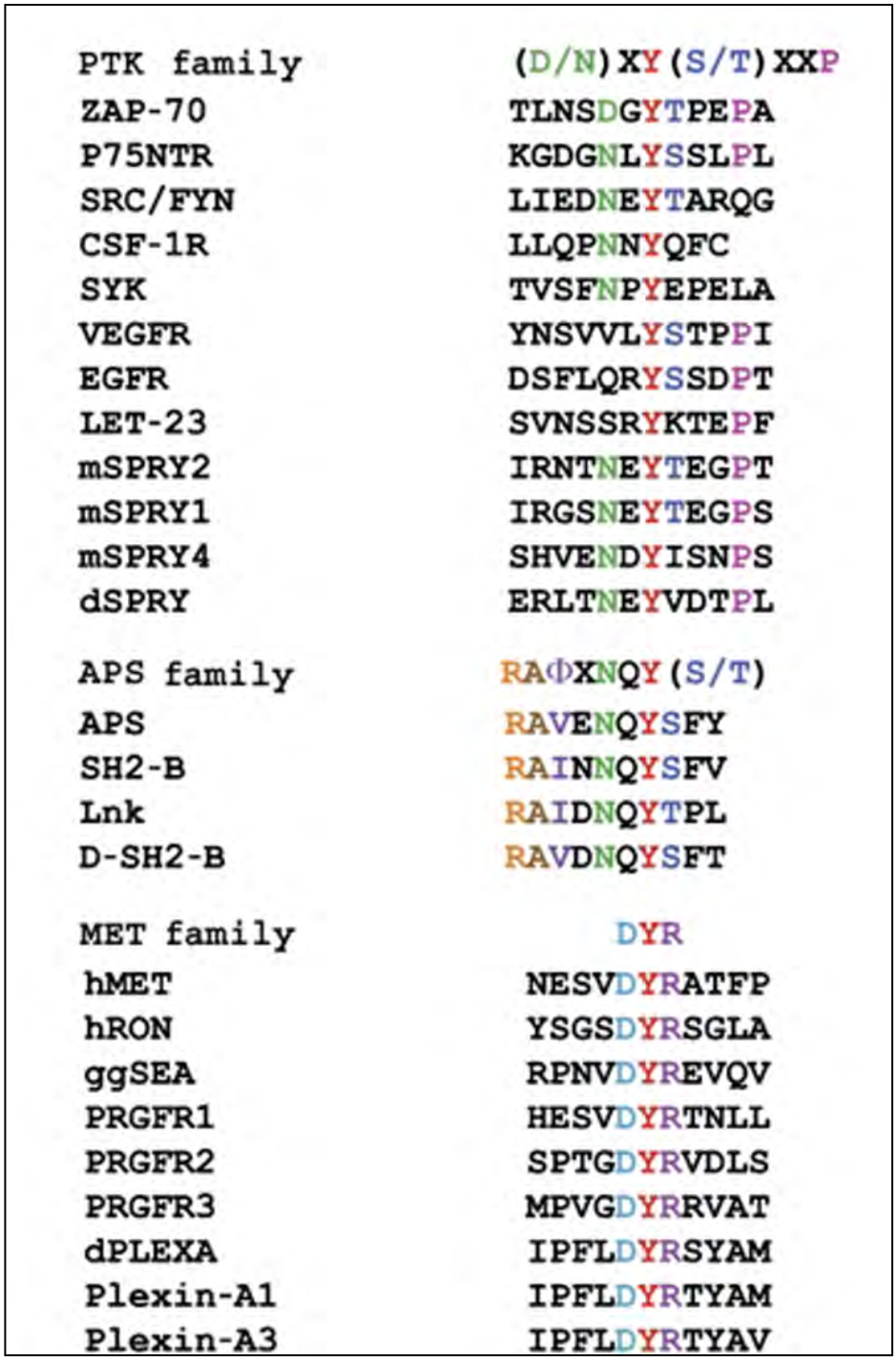



**Figure 1.22** Diverse pTyr motifs of c-Cbl TKB. Sequence alignment of c-Cbl-TKB binding sites for the PTK, APS and Met subgroups of proteins. This figure is adapted from (Ng et al., 2008), with permission from Macmillan Publishers Ltd.

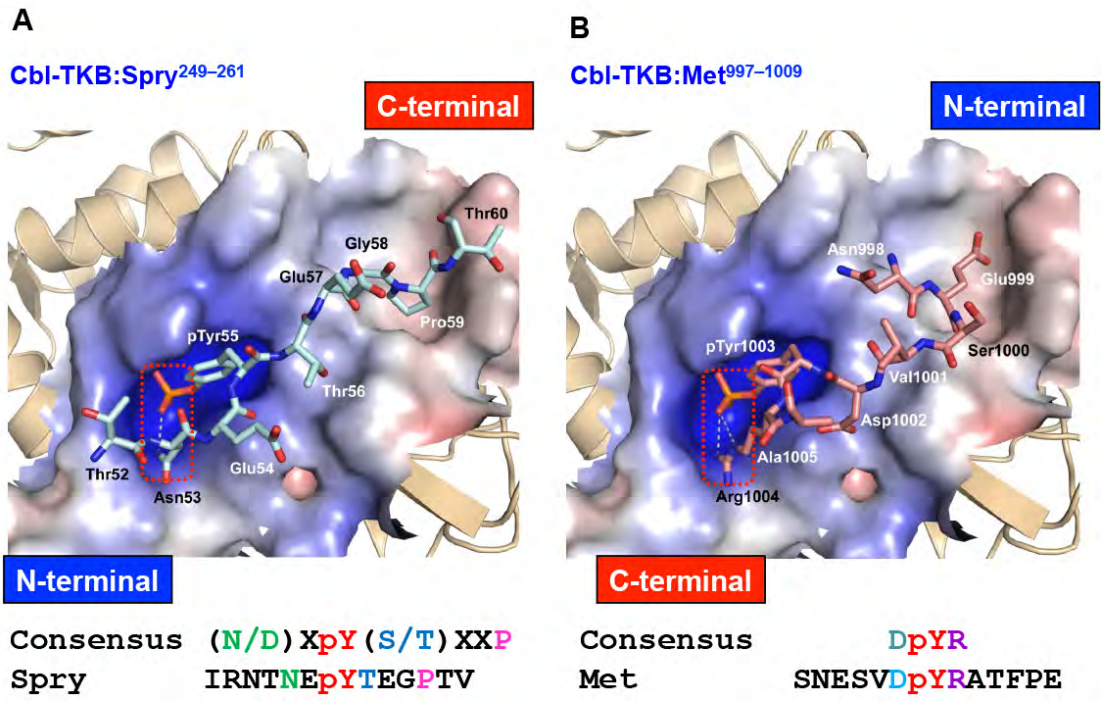



**Figure 1.23** Unique binding mechanism employed in c-Cbl TKB to recognize diverse pTyr motifs (A) Orientation of the intrapeptidyl hydrogen bond of the Spry$^{249–61}$ phosphopeptide within the binding pocket of the Cbl-TKB (B) Recognition of Met$^{997–1009}$ phosphopeptide involves the essential intrapeptidyl hydrogen bond and reverse binding of c-Cbl-TKB domain. This figure is adapted from (Ng et al., 2008), with permission from Macmillan Publishers Ltd.

### 1.8 Discovery of Hakai- a 'c-Cbl like' protein

E-cadherins play major role in epithelial cell adhesion, through the establishment of calcium dependent homophilic interactions at sites of cell-to-cell contact (Fujita et al., 2002; Pece and Gutkind, 2002). During tumor progression, disruption of cell-cell contacts plays a pivotal role in metastasis as well as in the dedifferentiation process that accompanies the malignant phenotype (Thiery and Sleeman, 2006). The most common malignancies in humans are carcinomas that are comprised of cells epithelial in origin (Guarino, 2010) and thus the loss of E-cadherin is considered as a major hallmark of tumor malignancy (Hanahan and Weinberg, 2000). The tyrosine kinase Src plays an active function in this process, as it phosphorylates tyrosine residues in the short intra-cytoplasmic tail of E-cadherins, thereby promoting their internalization

by endocytosis (Figure 1.24) (Fujita et al., 2002; Guarino, 2010; Pece and Gutkind, 2002). In search of the underlying molecular mechanism, Fujita et al. (2002) identified a new molecule, named Hakai (Hakai means 'destruction' in Japanese), that binds tyrosine phosphorylated E-cadherins and mediates the internalization and subsequent ubiquitin-dependent degradation of E-cadherins (Figure 1.24). Pece and Gutkind (2002) proposed a dynamic model for the function of Hakai mediated ubiquitination in E-cadherin endocytosis and trafficking, in which E-cadherins are internalized and either recycled to the plasma membrane or degraded in lysosomes (Fig. 1.24).

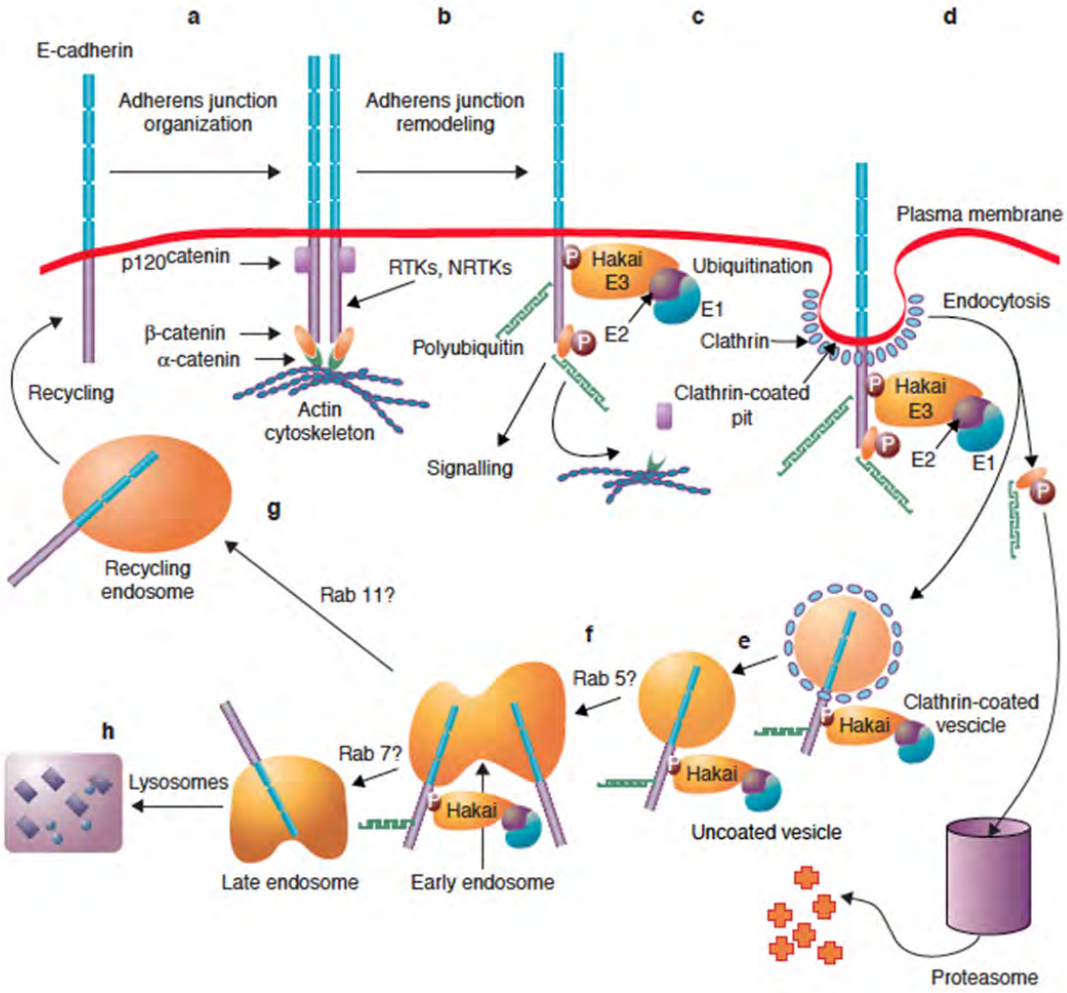



**Figure 1.24** A model for Hakai in the dynamic regulation of E-cadherin- mediated

adherens junction. **(a)** At steady-state, E-cadherins are organized in multiprotein complexes in adherens junctions. **(b)** Activation of receptor (RTKs) and non receptor (NRTKs) tyrosine kinases promotes the phosphorylation of β-catenin and the intracytoplasmic tail of E-cadherin (P), the recruitment of Hakai in a tyrosine-phosphoylation dependent manner, the loss of E-cadherin interaction with the actin cytoskeleton, and the disruption of adherens junctions. **(c)** E-cadherin bound to Hakai may initiate the activation of intracellular signalling pathways, whereas the E3-ligase function of Hakai mediates the transfer of ubiquitin chains to E-cadherin and β- atenin through the E1–E2 ubiquitination system. **(d)** Ubiquitinated β-catenin is degraded in the proteasome, but ubiquitinated E- adherin–Hakai complexes are likely internalized by clathrin-coated pits, which are then **(e)** rapidly uncoated and fuse to early endosomes. **(f)** In endosomes, depending on a balance between ubiquitination and de-ubiquitination systems, E-cadherin may be either recycled back to the cell surface **(g),** or targeted for degradation in the lysosomal compartment **(h)**. The possible function of members of the Rab family of GTPases in the intracellular trafficking of E-cadherin is depicted. This figure is adapted form (Pece and Gutkind, 2002), with permission from Macmillan Publishers Ltd.

Fujita et al. (2002) showed that Hakai functions as an E3 ubiquitin ligase that binds E-cadherin in a tyrosine phosphorylation-dependent manner. E3 ubiquitin ligases exhibit distinct structural features that determine their binding specificity and their ability to engage the ubiquitin conjugating enzymes. The latter includes the HECT domains, the RING finger domains, and the recently identified U-box domain. Hakai belongs to the second group, as it has a RING finger domain related to that of the c-Cbl. The relationship with c-Cbl goes beyond this domain and its function, as Hakai also exhibits a phosphotyrosine-binding domain and an extended proline-rich region that are highly related to those of c-Cbl (Figure 1.25). Furthermore, like c-Cbl, Hakai also binds to its target in a phosphotyrosine dependent manner resulting in its ubiquitination and subsequent degradation. Due to its similarity with c-Cbl, Hakai is also designated as CBLL1 (c-Cbl-like protein 1).

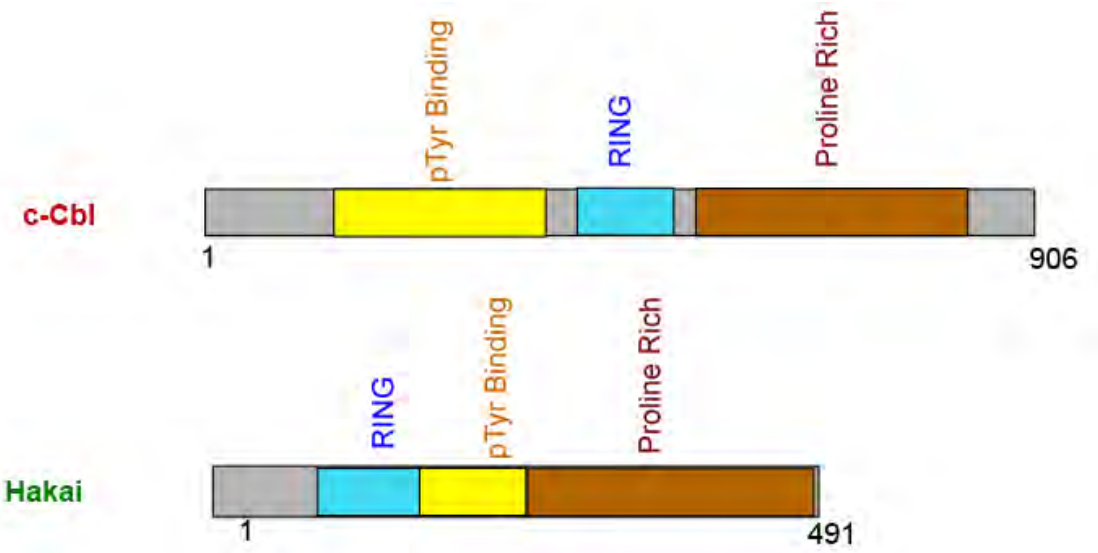



**Figure 1.25** Schematic diagrams of the Hakai and c-Cbl domains. There is a mild degree of sequence homology between Hakai and c-Cbl's RING finger (~20%) pTyr binding domain (~8%) and proline-rich region (~14%); however the domains are ordered differently on the respective proteins.

Hakai was originally identified as a key regulator of cell-cell contacts by controlling the fate of E-cadherin. Several lines of research over the last 10 years since the discovery of Hakai have delineated its emerging and crucial role in various cellular processes including adhesion, migration, embryogenesis, tumorigenesis and other diseases (Aparicio et al., 2012). A brief account of the knowledge that has been accumulated since Hakai discovery and its implication in human cancer and infectious disease is discussed in the following sections.

## 1.9 Important biological implications of Hakai

### 1.9.1 Hakai and Cancer

There are several lines of evidences that support multiple roles of Hakai in tumorigenesis. For instance, Hakai is highly up-regulated in human colon and gastric cancers, suggesting its direct involvement in cancer (Figueroa et al., 2009). Moreover, Hakai reduces cell attachment to the substrate and enhances the invasion capacity in

epithelial MDCK cells (Rodriguez-Rigueiro et al., 2011a). Furthermore, the overexpression of Hakai results in the down-regulation of Paxillin- a key focal adhesion associated protein, leading to the loss of cell adhesion in a proteasome-independent mechanism (Rodriguez-Rigueiro et al., 2011b). These observations substantiate the significant link between Hakai and cancer.

Considering the important role of Hakai in tumor progression, it could be an attractive target for therapeutic interventions against cancer.

### 1.9.2 Hakai and Infectious disease

There are several reports that describe Hakai's role in infectious disease. The first work describing Hakai in this process was by Krishnan et al. (Krishnan et al., 2008). By using a human genome-wide RNAi screen they identified Hakai as a protein that affect West Nile virus (WNV) infection. WNV belongs to the genus *flavivirus,* which constitute a significant global human health problem (Brinton, 2002). They proposed Hakai to be involved in the cellular internalization of WNV. They demonstrated that Hakai and the proteasome–ubiquitin systems are required for the cellular internalization of WNV (Krishnan et al., 2008). Still, the implication of Hakai during WNV infectious needs to be further clarified.

Other important studies on Hakai in infectious disease were reported in *Listeria monocytogenes*. Upon infection, *Listeria* leads to a wide range of symptoms associated to listeriosis, such as gastroenteritis, fetoplacental, and central nervous system infections (Lecuit, 2005). Internalization of *L. monocytogenes* mainly occurs *via* two bacterial surface proteins: internalin-A (InIA) and internalin-B (InIB) that has

E-cadherin and Met as their respective major host-cell surface receptors (Mengaud et al., 1997; Shen et al., 2000). InIA interaction with E-cadherin activates β- and α-catenin-mediated signaling pathways involved in the formation of adherens junctions. The initial signals triggered by the interaction of InIA with E-cadherin enhance the internalization of E-cadherin by Src-mediated tyrosine phosphorylation of E-cadherin followed by its ubiquitination by ubiquitin-ligase Hakai (Bonazzi et al., 2008).

### 1.9.3 Other biological implications of Hakai

Apart from the functional role of Hakai in cancer and infectious diseases, Hakai plays important role in other cellular processes and diseases. In a microarray profiling, Hakai was found differentially expressed during erythroid differentiation of murine erythroleukemia cells, suggesting the possibility of considering Hakai as a marker for erythropoiesis (Heo et al., 2005). Moreover, Hakai expression could be novel gene marker for immune-suppression in mouse lymph node assay (Oshida et al., 2011).

## 1.10 Objectives

Hakai is involved in the regulation of cell adhesion, cell migration and embryogenesis (Figueroa et al., 2009; Fujita et al., 2002; Gong et al., 2010; Kaido et al., 2009). Among the reported protein interactions of Hakai, its association and ubiquitination of the tyrosine phosphorylated E-cadherin upon Src activation is the best characterized (Fujita et al., 2002). The involvement of Hakai in modulating the E-cadherin mediated cell adhesion suggests that it participates in cancerous metastasis by disrupting the cell-cell contacts. Knowledge of the structure and function of Hakai, particularly the pTyr-binding domain of Hakai which binds the tyrosine phosphorylated E-cadherin

motif is crucial for understanding the whole process, as well as for identifying other potential targets of Hakai.

Our earlier studies on the pTyr-binding domain of c-Cbl (Ng et al., 2008; Sun et al., 2010; Sun et al., 2011) led us to examine the nature of the Hakai pTyr-binding domain to provide structural insights into its interaction with E-cadherin. Based on the sequence analysis, the Hakai pTyr-binding domain does not bear significant sequence homology with the known SH2 and PTB domains. Particularly, the putative pTyr-binding domain of Hakai bears less than 8% sequence identity with the pTyr-binding domain of c-Cbl. Furthermore, based on sequence analysis, the Hakai pTyr-binding domain comprised of around 50 residues, which is significantly smaller than all the existing pTyr-binding domains including SH2 and PTB domains. In addition, the pTyr motif of E-cadherin that interacts with the pTyr-binding domain of Hakai appears to be very unique as it contains three consecutive Tyrosine residues (Figure 1.26). Based on the uniqueness of the pTyr-binding domain of Hakai and the distinctive E-cadherin motif it recognizes, we hypothesize that a completely novel pTyr-binding domain is present in Hakai.

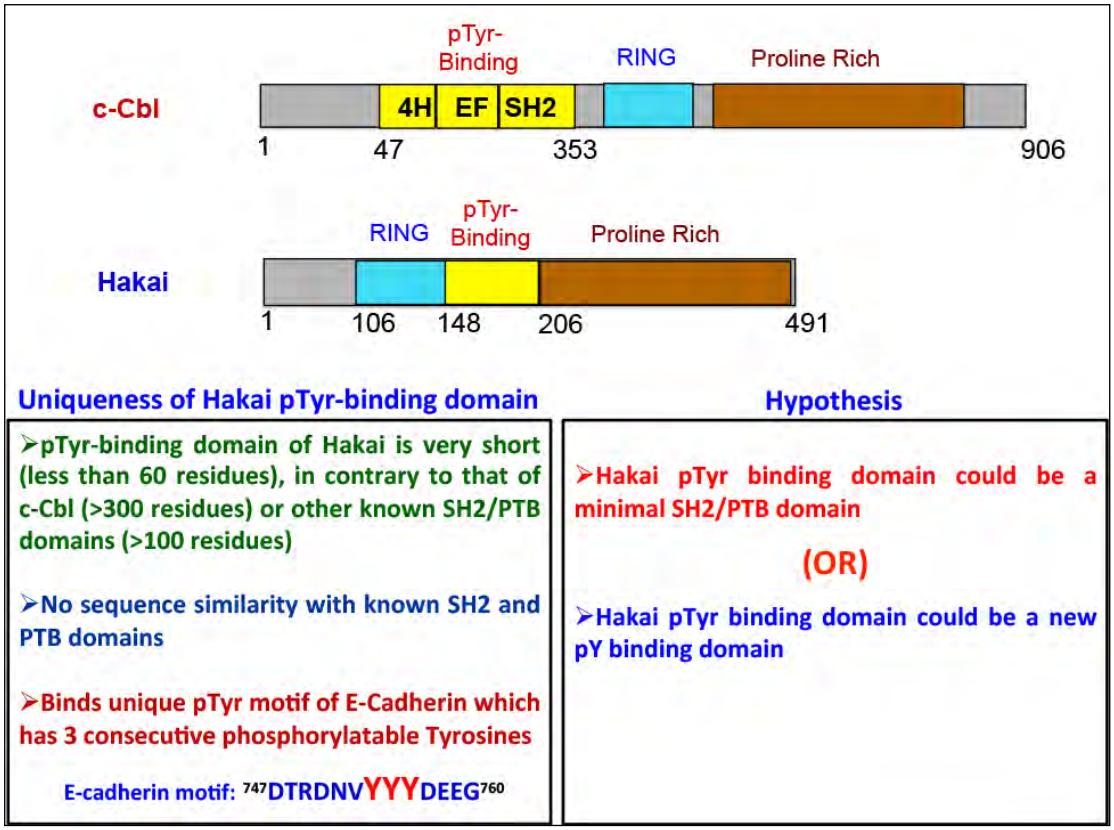



**Figure 1.26** Unique features of Hakai pTyr-binding domain. Although Hakai share functional similarity with c-Cbl, its pTyr-binding domain appears to be completely different. Moreover, Hakai pTyr-binding domain bears no similarity with the existing SH2 and PTB domains.

Thus, considering the unique features of Hakai pTyr-binding domain, the objectives of the present study are as follows:

1. To identify and characterize the domain boundaries that defines the pTyr-binding domain of Hakai
2. To characterize and determine the structure of the Hakai pTyr-binding domain. This will delineate the exact nature of this unique domain, whether it is a minimal SH2 or PTB domain, or a novel pTyr-binding domain altogether
3. To determine how Hakai pTyr-binding domain binds the pTyr motif of E-cadherin and validate the binding with point-mutational analyses using full-length proteins

4. To experimentally determine the binding partners of Hakai pTyr-binding domain and characterize their interaction with Hakai pTyr-binding domain, and validate the findings using full length proteins
5. To determine if Hakai like pTyr-binding exist in other human proteins by sequence based studies and experimental approaches
6. To carry out NMR and other solution studies to gain better understanding about the additional attributes of the Hakai pTyr-binding domain

# CHAPTER II

## Structure and function of a novel phosphotyrosine-binding domain in Hakai- the HYB domain

## 2.1 Introduction

In eukaryotic cells, phosphorylation events regulate cell signalling by providing docking sites for protein domains, such as the Src homology 2 (SH2) and phosphotyrosine binding (PTB) domains (Forman-Kay and Pawson, 1999; Pawson and Nash, 2003; Yaffe, 2002). The SH2 was the first signalling domain to be identified and has been extensively characterized (Filippakopoulos et al., 2009; Forman-Kay and Pawson, 1999; Liu et al., 2006; Pawson and Nash, 2003; Songyang et al., 1993; Yaffe, 2002). The SH2 is a dedicated phosphotyrosine-binding domain and plays a critical role in signal transduction, hence making it a target for drug development (Kasembeli et al., 2009; Machida and Mayer, 2005a; Pawson and Nash, 2003; Taylor et al., 2008). Binding specificity of SH2 domains is generally conferred by the sequences flanking the C-terminus of the phosphotyrosine (pTyr), and motif recognition is usually relatively inflexible. The other major class of pTyr-binding domain is the PTB domain. The specificity of binding to the PTB domain is conferred typically by residues on the target that are N-terminal to the pTyr. However, the PTB domain also recognizes non-pTyr motifs (Farooq and Zhou, 2004; Forman-Kay and Pawson, 1999; Pawson and Nash, 2003; Smith et al., 2006; Yaffe, 2002). Atypical phosphotyrosine-binding domains have also been detected in PKCδ and the human M2 pyruvate kinase (PKM2) (Benes et al., 2005; Christofk et al., 2008).

In 2002, Fujita et al (2002) discovered a novel ubiquitin E3 ligase protein that targeted pTyr sites on E-cadherin. The E3 ligase, Hakai, possesses three domains: a RING domain, a short pTyr recognition sequence and a proline-rich domain (Fujita et al., 2002). Hakai is involved in the regulation of cell adhesion, cell migration and embryogenesis (Figueroa et al., 2009; Gong et al., 2010; Kaido et al., 2009). Among

the reported protein interactions of Hakai, its association with and ubiquitination of E-cadherin upon Src activation is the best characterized (Fujita et al., 2002).

Based on molecular modeling, Fujita et al (2002) assumed the pTyr-binding domain of Hakai to be a derivative SH2 domain. Our previous experience analyzing the SH2 domain of a similar E3 ubiquitin ligase, c-Cbl (Ng et al., 2008) led us to examine the nature of the Hakai pTyr-binding domain to provide structural insights into its interaction with E-cadherin. In this study, we report that the Hakai pTyr-binding domain consists of a homodimer formed at a structurally novel interface. Each monomer consists of two zinc-finger domains: a RING domain and a minimum pTyr-binding domain that incorporates a novel, atypical zinc coordination motif. Both domains play key roles in dimerization. The Hakai pTyr-binding domain is therefore composed of four zinc-binding domains cooperating to bind pTyr residues surrounded by acidic amino acids. Whereas the RING domain appears in other proteins, the atypical zinc-binding domain component is a novel protein fold that incorporates an intertwined configuration.

The work reported in this chapter was published in EMBO J (31:5:p 1055 – 1329). We have collaborated with Dr Guy Graeme, IMCB Singapore to conduct the cell-based experiments in his lab. Dr Chow Soah Yee and Dr Permeen Yusoff are mainly involved in these experiments from Dr Guy's lab.

## 2.2 Materials and Methods

### 2.2.1 Plasmids

For structural studies, Hakai constructs were cloned into a set of vectors for construct optimization. These include pGEX-6P-1 (GE Healthcare, UK), pGEX-4T-1 (GE Healthcare, UK), pET-32a (Novagen, USA), pET-M (modified pET-32a constructed

by removing the Trx tag and S tag), pET SUMO (Invitrogen, USA), pET-44a (Novagen, USA) and pET PPL His6 MBP LIC cloning vector (Addgene, Cambridge, MA, plasmid 37183, deposited by Scott Gradia). The final optimized constructs for structural studies were those in pGEX-6P-1 vector inserts corresponding to mouse Hakai (148aa-206aa) and Hakai (106aa-206aa), respectively. Point mutations for the constructs used in structural and biophysical studies were made using KAPA HiFi DNA Polymerase (Kapa Biosystems, USA).

Mouse Hakai, human E-cadherin and avian v-Src were gifts from W. Birchmeier (Max- Delbrück-Center for Molecular Medicine, Germany), W. Hunziker (IMCB, Singapore) and X.M. Cao (IMCB, Singapore), respectively. Mouse cortactin was from Addgene (Cambridge, MA) (plasmid 26722, deposited by A. Weaver). Human ZNF645 and DOK1 were from Origene (Rockville, MD). Where necessary, the genes were cloned into pXJ40-HA or pXJ40-FLAG. Point mutants and truncates for mammalian cell based studies were generated using the proofreading *Pfu* DNA polymerase.

### 2.2.2 Antibodies and reagents

Mouse anti-FLAG M2, rabbit anti-FLAG and anti-HA, and agarose-conjugated anti-FLAG M2 beads were from Sigma-Aldrich (St. Louis, MO). Rabbit GST, cortactin, E-cadherin and DOK1 antibodies were purchased from Santa Cruz Biotechnology (Santa Cruz, CA). Protein A-conjugated agarose beads were from Roche Molecular Biochemicals (Germany). Mouse anti- E-cadherin and HRP-conjugated anti-phospho-tyrosine PY20 were from BD Transduction Laboratories (Lexington, KY). Mouse

anti-β-actin was from Abcam (Cambridge, MA).

### 2.2.3 Cell lines and transfection

HEK293 cells were purchased from ATCC (Manassas, VA) and maintained as described (Yusoff et al., 2002). Transfections were performed using Lipofectamine 2000 (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions.

### 2.2.4 Immunoprecipitation and immunoblotting

Immunoprecipitation and immunoblotting were carried out as described (Yusoff et al., 2002) with the following modifications. HEK293 cells were harvested 24 h post-transfection with a lysis buffer containing protease inhibitors (Roche) and 1mM $Na_3VO_4$ ($Na_3VO_4$ is an inhibitor of protein tyrosine phosphatases, alkaline phosphatases and ATPases). Immunoprecipitations were performed using agarose-conjugated anti-FLAG or protein-specific antibodies followed by incubation with protein-A-conjugated agarose beads at 4°C.

### 2.2.5 Protein expression and purification

Inserts corresponding to Hakai (aa 148-206) and Hakai (aa 106-206) were cloned into pGEX-6P-1 vector (GE Healthcare) using *BamHI* and *SalI* as the restriction sites and expressed as a GST-fusion protein in *E. coli* BL21 (DE3) cells. The cells were cultured in Luria Broth medium at 37°C until the $A_{600}$ nm reached 0.6–0.7, IPTG and zinc sulphate were then added at concentrations of 0.15mM and 50μM, respectively, and growth continued for around 20 hours at 15°C. Cells were harvested by centrifugation (8000 × g, 10 min, 4°C), and the pellet was resuspended in lysis buffer

(50 mM Bis-Tris pH 6.5, 300mM NaCl, 10μM zinc chloride, 5% Glycerol, 0.5% Triton-X 100, 2mM DTT, 1mM phenylmethylsulfonylfluoride (PMSF) and homogenized using a French Press Cell Disrupter (Thermo Electron Corporation). Subsequently, the cell lysate was centrifuged at 18,000 rpm for 30 min at 4°C (JA-25.50 fixed angle rotor, Beckman Coulter centrifuge) and the supernatant was allowed to bind to glutathione-sepharose resin (GE Healthcare) for 2-4 hours at 4°C and was subsequently washed with a buffer containing 50 mM Bis-Tris pH 6.5, 300mM NaCl, 5% glycerol, 2mM DTT. After the washing step, an overnight on-column cleavage at 4°C with GST-PreScission Protease (GE Healthcare) was performed to remove the GST tag. A major portion of GST and GST-PreScission Protease remained bound to the glutathione-sepharose resin and the flow-through containing the partially purified, untagged proteins were further purified using a Superdex 75 size exclusion column (GE Healthcare) equilibrated with a buffer containing 10 mM Bis-Tris pH 6.5, 250mM NaCl and 5mM DTT.

### 2.2.6 NMR spectroscopy and chemical shift perturbation analysis

$^{15}N/^{13}C$-labelled Hakai (aa 106–206) was obtained from cultures grown in M9 media supplemented with $^{15}N$-labelled ammonium chloride and $^{13}C$-labelled glucose as the sole nitrogen and carbon sources, respectively. The labeled proteins were purified as described above. NMR spectra were acquired at 298 K in an 800-MHz NMR spectrometer (Bruker, Karlsruhe, DE). The backbone assignment was obtained using standard $^{15}N$-edited HSQC, HNCACB and CBCA (CO)NH experiments; side chains were assigned using standard 3D-TOCSY, 3D-NOESY and HCCH-TOCSY experiments. NMR data were processed using NMRPipe (Delaglio et al., 1995) and analysed by NMRView (Johnson and Blevins, 1994).

For the chemical shift perturbation analysis, the 2D $^{1}H–^{15}N$-HSQC spectra for the $^{15}N$-labelled Hakai (aa 106–206) were acquired in the absence or presence of the phosphorylated E-cadherin peptide. Perturbed residues on Hakai were assigned by superimposing the two HSQC spectra.

### 2.2.7 Crystallization and structure determination

SelMet-substituted Hakai (aa 106-206) was expressed in a methionine auxotroph (Doublie, 1997) and purified as described above. SelMet incorporation was verified by MALDI-TOF.

SelMet Hakai (aa 106–206) crystals were grown at 289 K by the hanging drop vapour diffusion method. The protein (30 mg/ml) was mixed with an equal volume of reservoir solution (140mM $Li_2SO_4$; 100mM Tris, pH 7.8; 15mM $Na_2S_2O_3$; 20–22% PEG 5000 MME; 1% isopropanol and 20–25% ethylene glycol). A complete SAD data set was collected to 1.9Å resolution at the synchrotron beamlines (NSLS, Brookhaven National Laboratory and the National Synchrotron Radiation Research Center [NSRRC], Taiwan) using a Quantum4-CCD detector (Area Detector Systems Corp., Poway, CA). The data sets were processed and scaled using HKL2000 (Otwinowski and Minor, 1997). The crystals belonged to space group $P6_222$ with a=64.66Å, b=64.66Å, c=121.04Å, and contained one molecule in the asymmetric unit.

All four expected selenium sites in the asymmetric unit were located by SOLVE

(Terwilliger and Berendzen, 1999). Initial phases were developed by RESOLVE (Terwilliger, 2003); the overall figure of merit was improved to 0.83, and over 90% of the molecule was built automatically. The remaining parts of the model were built manually using COOT (Emsley and Cowtan, 2004) and alternatively refined by CNS (Brunger et al., 1998) and PHENIX (Adams et al., 2002). The final model was refined to a 1.9-A° resolution with an R-factor of 0.2175 ($R_{free}$=0.2396) and analysed using PROCHECK (Laskowski et al., 1993). All structure-related figures were prepared using PyMOL (DeLano, 2002).

### 2.2.8 Circular Dichroism spectrometry

Far UV spectra (260–190 nm) of Hakai (aa 148–206) and Hakai (aa 106–206) and its mutants were measured using a Jasco J-810 spectropolarimeter in phosphate buffer (pH 7.5) at room temperature using a 0.1-cm path length and stoppered cuvettes. Six scans were recorded, averaged and the baseline subtracted.

### 2.2.9 Isothermal Titration Calorimetry

Phosphorylated and non-phosphorylated peptides of E-cadherin (residues 749–761); phosphorylated peptides of DOK1 (residues 356-366) and Cortactin (residues 477-489) were titrated at a molar concentration of 800 μM against 100 μM of Hakai (aa 106– 206) dimer in a VP-ITC microcalorimeter (Microcal, Northhampton, UK) at 293 K. 30 titrations were carried out using 10-μl injections of the appropriate peptide into the sample cell containing Hakai (aa 106–206) and the data were analysed with a one-site binding model using the Origin software package v7.0 supplied by Microcal. All measurements were repeated thrice.

### 2.2.10 Dynamic light scattering

Dynamic light scattering studies were carried out on a DynaPro Light Scattering instrument (Protein Solutions, USA) at a protein concentration of 2 mg/ml, in a buffer containing 10mM Bis-Tris pH 6.5, 250mM NaCl and 5mM DTT.

### 2.2.11 Liquid chromatography–mass spectrometry/mass spectrometry

Immunoprecipitates were separated by SDS–PAGE, and stained with Coomassie Blue G250. Protein bands were excised and washed with 25mM ammonium bicarbonate (ABB) in 50% acetonitrile (ACN) buffer thrice. The proteins in the gel were reduced with 10mM DTT in 25mM ABB buffer, alkylated with 5mM iodoacetamide, dehydrated and digested with trypsin overnight. After in-gel digestion, the solution was transferred to a clean tube and sonicated for 30 min in the presence of 50 ml 50% ACN and 5% acetic acid for protein extraction. This extraction procedure was repeated three times; the pooled extracts were dried with a vacuum concentrator. The samples were processed and analysed as described (Zhang et al., 2010) using a LTQ-FT ultra mass spectrometer (Thermo).

For each experiment, MS/MS (dta) spectra were extracted from the raw data files using the extract_msn program in Biowork 3.3 (ThermoFinnigan). The extracted dta files were combined into a single file in the Mascot generic file (mgf) format. Except for the conversion of precursor mass from MHþ in dta to *m/z* in mgf, the fragment ion *m/z* and intensity values were used as determined. Proteins were identified by searching the combined data against the IPI human database (downloaded on 25

November 2009, including 86845 sequences and 35122 444 residues) *via* an in-house Mascot server (version 2.2.07). Two missing cleavages were allowed. Precursor ion and MS/MS fragment ion error tolerances were set to <10 ppm. and <20 ppm, respectively. A protein was accepted as a true positive if it had a significant score ($P<0.05$) and at least two unique peptides.

### 2.2.12 PDB Accession Code

The coordinates and structure factors have been deposited at the Protein Data Bank (PDB) with the accession code 3VK6.

## 2.3 Results

### 2.3.1 Determination of the minimum pTyr- binding domain in Hakai that interacts with tyrosine phosphorylated E-cadherin

To establish the minimum region on Hakai necessary to bind to E-Cadherin we first reiterated the binding assay utilized by Fujita et al., 2002. From Figure 2.1A it can be seen that E-Cadherin binds to Hakai only when tyrosine phosphorylated by v-Src. In a number of later experiments we will only show the v-Src stimulated samples for clarity of presentation.

To identify the region of interaction between E-cadherin and Hakai, we made a number of Hakai constructs (as indicated in Figure 2.1B and Figure 2.1D) and tested their ability to bind to overexpressed E-Cadherin after phosphorylation by v-Src. Experiments using these constructs are shown in Figure 2.1C and Figure 2.1E respectively, and the results indicated that the designated minimum pTyr-binding

domain consisted of amino acids 148-206, which is considerably smaller than the approximately 100 amino acid SH2 or PTB domain.

In an additional experiment, we made a few more constructs of Hakai as indicated in Figure 2.1F to further confirm the minimum pTyr-binding requirement. Here we made an interesting observation and demonstrated that binding was stronger when the RING domain was also included in binding studies (Figure 2.1G right panel). It is notable that we had to attach GST to the 148-206 sequence of Hakai in order to stably express it. This indicates that Hakai (148-206) might require additional structural elements and stabilization factors to achieve a stable fold.

Thus, the combined results demonstrates that Hakai (aa148-206) is the minimum pTyr-binding domain while Hakai (aa106-206) which includes the RING domain in addition to the minimum pTyr-binding domain binds stronger with tyrosine phosphorylated E-cadherin.

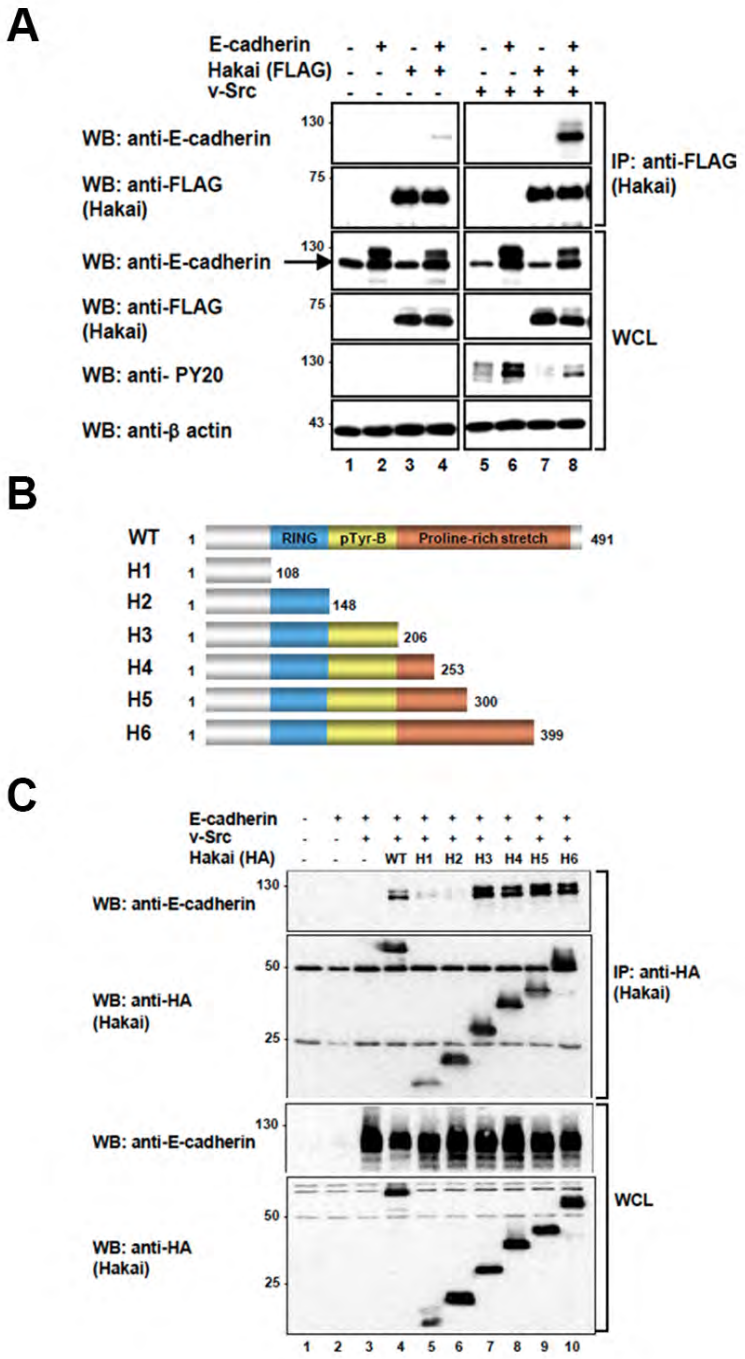



**Figure 2.1 (A-C)** Determining the minimum-binding region of Hakai. **(A)** Wild-type E-cadherin and Hakai were overexpressed in HEK293 cells in the presence or absence of v-Src. The arrow indicates a non-specific band. E-cadherin is separated into two bands by SDS-PAGE, and the lower band co-migrates with the non-specific band. **(B)** A schematic representation showing the domains and truncation constructs (H1–H6) of Hakai: RING (RING domain), pTyr-B (phosphotyrosine-binding domain), Pro-rich stretch (Proline-rich domain). **(C)** E-cadherin was co-transfected into HEK293 cells with the HA-tagged wild-type or truncation constructs of Hakai indicated in (B) to determine which section of the C-terminal region of Hakai is required for interaction with E-cadherin.

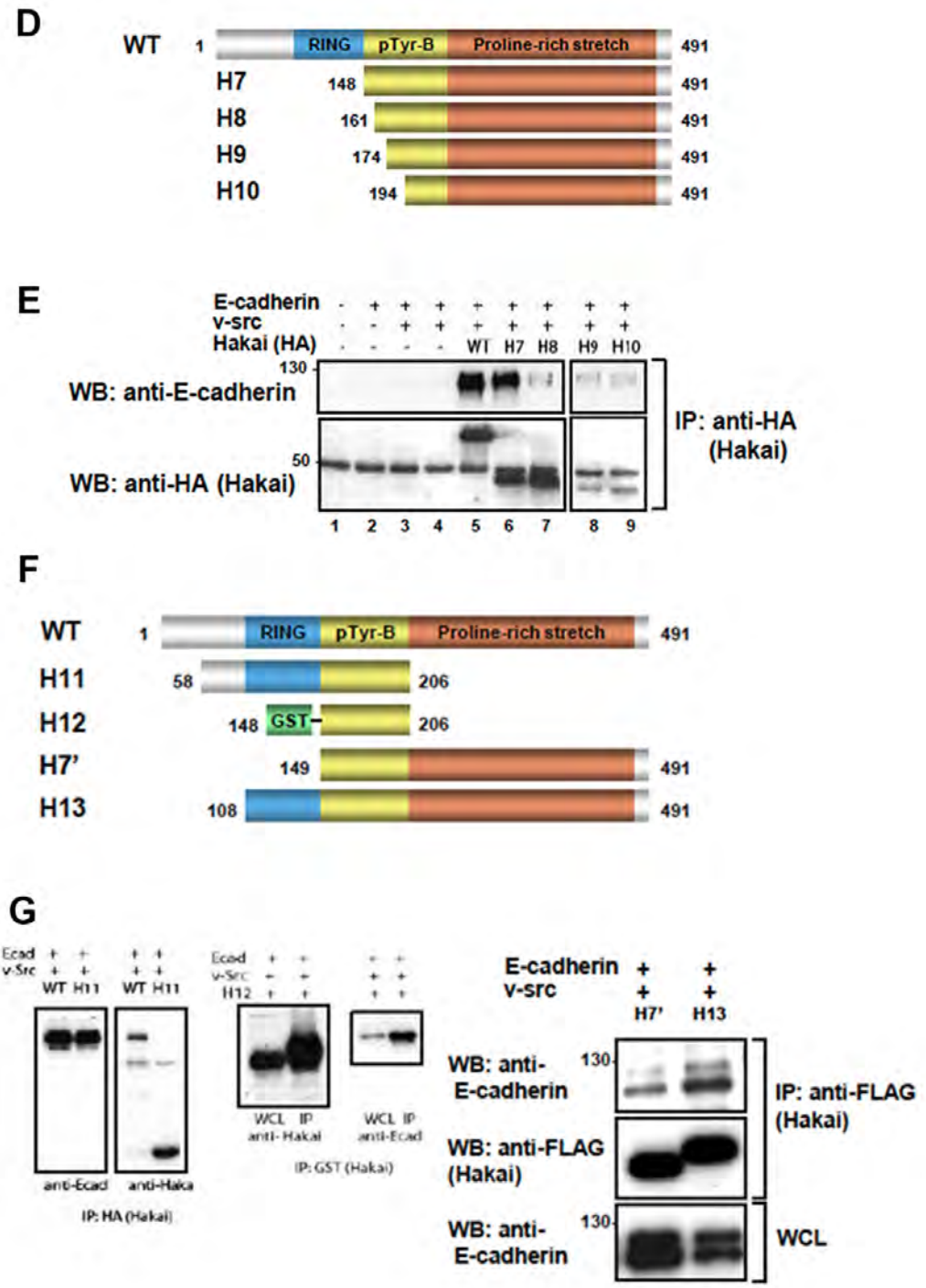



**Figure 2.1 (D-G)** Determining the minimum-binding region of Hakai (*continued*). **(D)** A schematic diagram showing the truncation constructs of Hakai (H7–H10). **(E)** HEK293 cells were co-transfected with E-cadherin plus HA-tagged wild-type Hakai or truncation constructs H7 to H10 to determine which section of the N-terminal region of Hakai is needed for interaction with E-cadherin. **(F)** A schematic representation of the Hakai constructs H11, H12, H7' and H13. Construct H12 was cloned into the mammalian expression vector pXJ40GST. **(G)** Wild-type Hakai and the truncation constructs H11, H12, H7' and H13 were co-transfected into HEK293 cells with E-cadherin to determine the region of Hakai to which E-cadherin binds.

### 2.3.2 Construct optimization

We experimentally identified that Hakai (aa 148-206) is the minimum pTyr-binding domain of Hakai. We also established that Hakai (aa 106-206), which contained both the RING domain as well as the minimum pTyr-binding sequence, binds stronger to tyrosine phosphorylated E-cadherin than the minimum pTyr-binding domain alone. Subsequently, we sought to gain structural and functional insights into the determinants of pTyr-binding present in Hakai. For this, we proceeded to express and purify both Hakai (aa148-206) as well as Hakai (aa106-206) in order to produce crystallization quality protein aimed for structure determination by employing X-ray crystallography.

In order to purify multi-milligram quantities of protein of crystallization quality, we made around 50 different constructs of Hakai containing a range of domain boundaries including Hakai (aa148-206) and Hakai (aa106-206) in a number of expression vectors, as described in the Materials and Methods section. Some of the construct details are summarized in Table 2.1. Finally, the Hakai constructs in pGEX-6P-1, that contains GST as solubility tag and a PreScission Protease cleavage site yielded stable protein, and the tag could be cleaved and separated by employing on-column PreScission Protease cleavage (Figure 2.2 and Figure 2.3). The construct details and results are summarized in Table 2.1. The expression and purification of Hakai constructs (aa 148-206) and (aa 106-206) in pGEX-6P-1 is explained as Materials and Methods. The final gel filtration profiles of Hakai (aa 148-206) and Hakai (aa 106-206), and the protein purity as seen in SDS-PAGE are shown in Figure 2.2 and Figure 2.3 respectively. The gel filtration profile of Hakai (aa 106-206)

indicated that it eluted as a dimer.

Hakai (aa 148-206) [minimum pTyr - binding region]
Hakai (aa 106-206) [RING + minimum pTyr - binding region]
Cloned in various vectors

| Vector | Solubility tag | Cleavage site | Protein Solubility | Result |
|---|---|---|---|---|
| pET-M | None(N-ter His tag) | None | Insoluble | Insoluble |
| pET-32a | Trx | Thrombin | Insoluble | Insoluble |
| pGEX-4T-1 | GST | Thrombin | Soluble | Tag could not be cleavage with thrombin |
| pET-44a | NusA | Thrombin | Soluble | NusA tag could not be cleaved |
| pET-44a | NusA | TEV protease site | Soluble | NusA tag could not be cleaved |
| pLIC-MBP | MBP | TEV protease site | Soluble | MBP tag could not be cleaved |
| Modified pET-28a | SUMO | TEV protease site | Soluble | Could cleave the SUMO tag with TEV protease |
| Modified pET-28a | SUMO | Cleaved using SUMO protease (ULP1) | Soluble | Tag was cleaved but could not be separated from the protein upon cleavage |
| pGEX-6P-1 | GST | PreScission Protease | Soluble | Tag was successfully cleaved as well as separated from the protein of interest |

**Table 2.1** Construct optimization for structural studies. Construct optimization of Hakai (aa148-206) and Hakai (aa106-206) aimed to produce multi-milligram quantities of the proteins of interest without tag for crystallization purpose

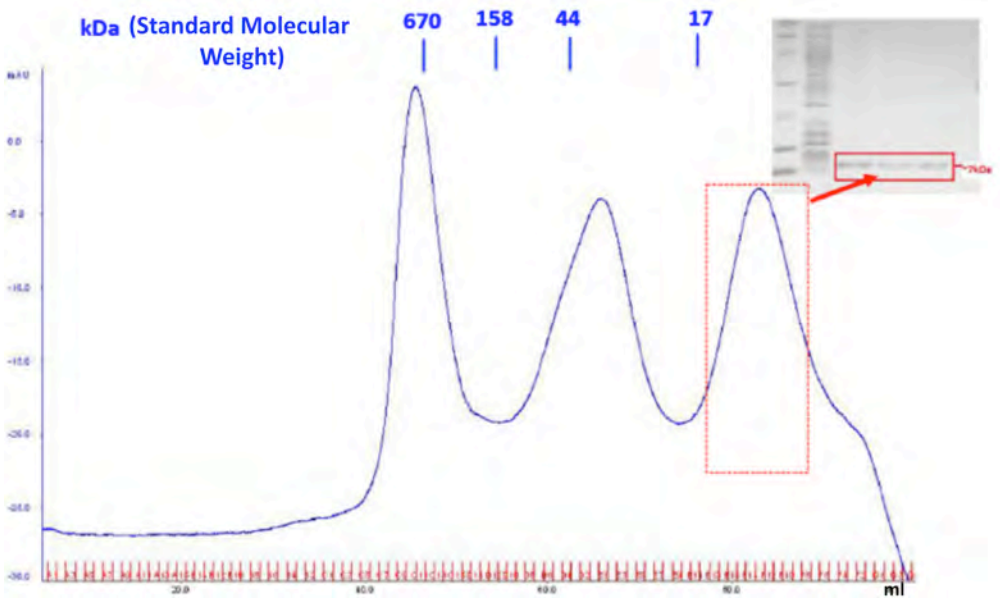



**Figure 2.2** Purification of Hakai (aa 148-206). Gel filtration profile of Hakai (aa 148-206) in Superdex-75 column. The protein was purified to homogeneity, as Hakai (aa 148-206) eluted as a single peak in the gel filtration. Gel filtration fractions from the peak were analyzed through 15% SDS-PAGE analysis. The X-axis indicates the elution volume in mL; Y-axis indicates the UV absorbance (at 280 nm) measured in mAU.

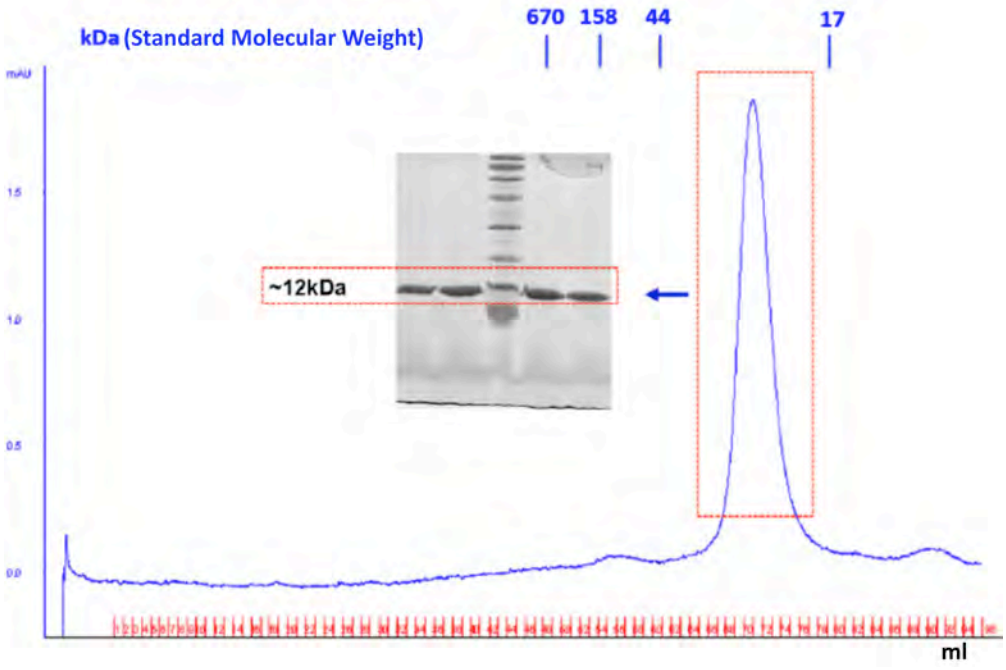



**Figure 2.3** Purification of Hakai (aa 106-206). Gel filtration profile of Hakai (aa 106-206) in Superdex-75 column. The protein was purified to homogeneity, as Hakai (aa 106-206) eluted as a single peak and the fractions from the peak were analyzed through 15% SDS-PAGE analysis. The X-axis indicates the elution volume in mL; Y-axis indicates the UV absorbance (at 280 nm) measured in mAU.

Upon obtaining the purified proteins for both Hakai constructs, *viz.* (aa 148-206) and Hakai (aa 106-206), we used Circular Dichroism (CD) experiment to analyze the folding and secondary structural elements of both the constructs. CD analysis however revealed that purified Hakai (aa 148–206) was unstructured (Figure 2.4), and, thus, was not suitable for crystallization. On the other hand, Hakai (aa 106-206), that contained both the RING domain and the minimum pTyr-binding domain was folded as revealed in the CD spectrum (Figure 2.4) and was well suited for the subsequent crystallization experiments.

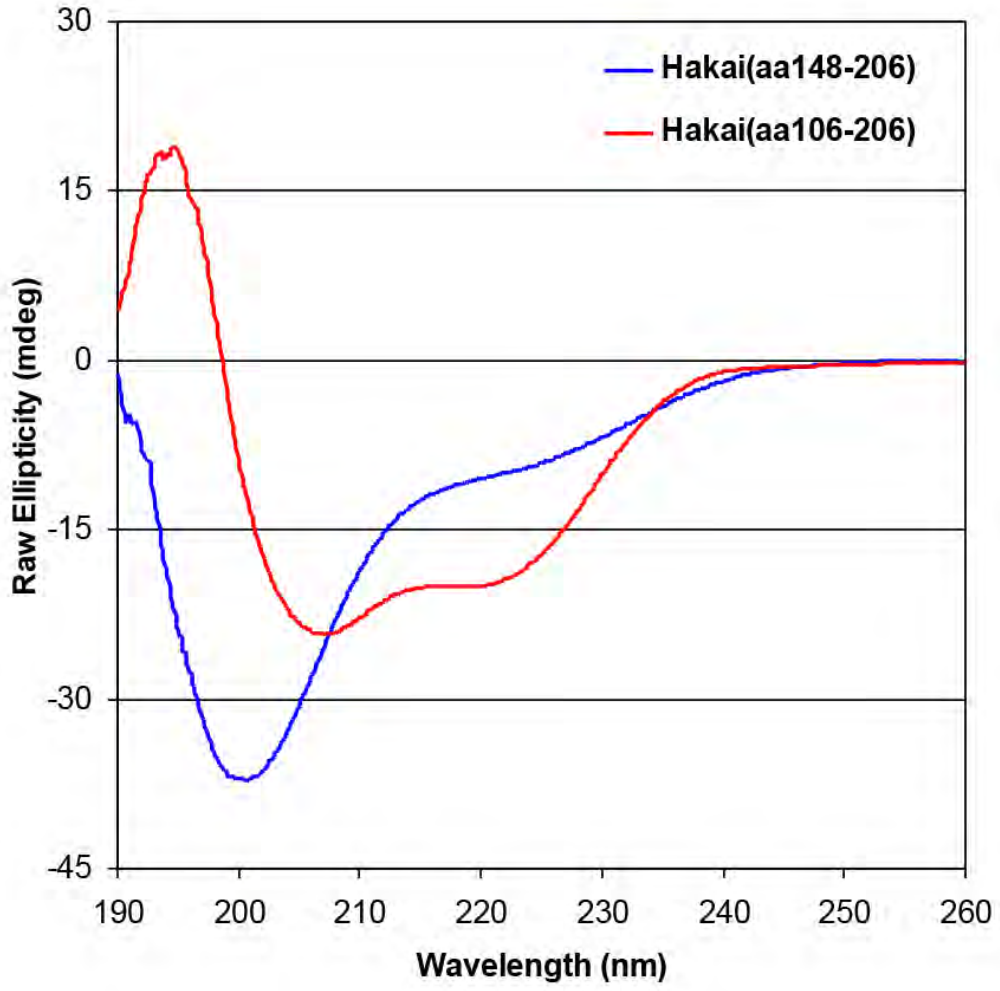
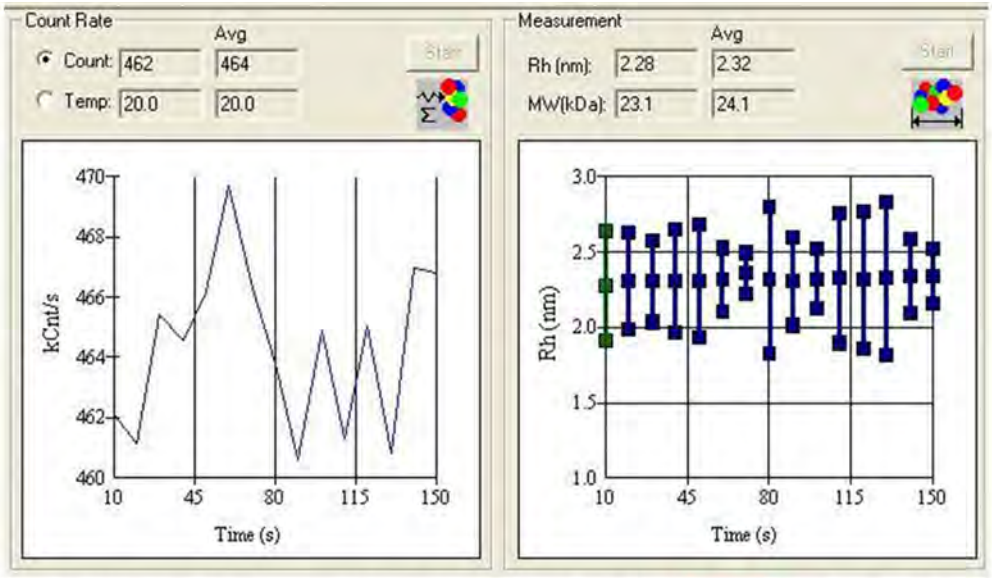



**Figure 2.4** CD spectra of Hakai (aa 148-206) and Hakai (aa 106-206). Readout from Circular Dichroism (CD) analysis showing that Hakai (aa 148-206) has no well defined secondary structure when compared to Hakai (aa 106-206).

Dynamic Light Scattering (DLS) experiment was performed during sample concentration of Hakai (aa 106-206) and prior to crystallization screening, to analyze the homogeneity of the protein. DLS experiment indicated a dispersity index of 0.02 at 30 mg/ml of protein, and this indicated that the Hakai (aa 106-206) was highly mono-dispersed and thus suitable for subsequent crystallization trials (Figure. 2.5). Also, the apparent molecular weight indicated by the DLS experiment suggested that Hakai (aa 106-206) existed as a dimer in solution.




| Mst# | Time(s) | Temp(C) | Count Rate | Ampl | Diff Coeff | Radius(nm) | Polyd(nm) | PolydIndx | MW(KDa) | %Mass | Baseline | Sos Error |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 10.0 | 20.0 | 462117 | 0.687 | 940. | 2.28 | 0.363 | 0.03 | 23.1 | 100.0 | 1.000 | 0.766 |
| 2 | 20.0 | 20.0 | 461132 | 0.680 | 928. | 2.31 | 0.321 | 0.02 | 23.8 | 100.0 | 1.000 | 0.850 |
| 3 | 30.0 | 20.0 | 465421 | 0.676 | 928. | 2.31 | 0.269 | 0.01 | 23.8 | 100.0 | 1.000 | 0.919 |
| 4 | 40.0 | 20.0 | 464574 | 0.683 | 928. | 2.31 | 0.344 | 0.02 | 23.8 | 100.0 | 1.000 | 1.20 |
| 5 | 50.0 | 20.0 | 466135 | 0.687 | 927. | 2.31 | 0.373 | 0.03 | 23.9 | 100.0 | 1.000 | 1.09 |
| 6 | 60.0 | 20.0 | 469686 | 0.682 | 924. | 2.32 | 0.212 | 0.01 | 24.0 | 100.0 | 1.000 | 0.923 |
| 7 | 70.0 | 20.0 | 466260 | 0.693 | 908. | 2.36 | 0.138 | 0.00 | 25.0 | 100.0 | 1.000 | 0.728 |
| 8 | 80.0 | 20.0 | 463812 | 0.691 | 925. | 2.31 | 0.486 | 0.04 | 24.0 | 100.0 | 1.000 | 1.04 |
| 9 | 90.0 | 20.0 | 460603 | 0.710 | 928. | 2.31 | 0.292 | 0.02 | 23.8 | 100.0 | 1.000 | 0.685 |
| 10 | 100.0 | 20.0 | 464878 | 0.700 | 922. | 2.32 | 0.197 | 0.01 | 24.2 | 100.0 | 1.000 | 0.611 |
| 11 | 110.0 | 20.0 | 461326 | 0.702 | 920. | 2.33 | 0.430 | 0.03 | 24.3 | 100.0 | 1.000 | 1.33 |
| 12 | 120.0 | 20.0 | 465031 | 0.718 | 925. | 2.31 | 0.454 | 0.04 | 24.0 | 100.0 | 1.000 | 1.30 |
| 13 | 130.0 | 20.0 | 460839 | 0.727 | 920. | 2.33 | 0.510 | 0.05 | 24.3 | 100.0 | 1.000 | 1.06 |
| 14 | 140.0 | 20.0 | 466947 | 0.727 | 915. | 2.34 | 0.245 | 0.01 | 24.6 | 100.0 | 1.000 | 1.46 |
| 15 | 150.0 | 20.0 | 466821 | 0.742 | 915. | 2.34 | 0.183 | 0.01 | 24.6 | 100.0 | 1.000 | 2.09 |
| Aves: | | | | | | | | | | | | |
| Mono | | 20.0 | 464372 | 0.700 | 923. | 2.32 | 0.321 | 0.02 | 24.1 | 100.0 | 1.000 | 1.07 |
| Bi-1 | | 0.0 | 0 | 0.000 | 0.000 | 0.000 | ---- | ---- | 0.000 | 0.0 | 0.000 | 0.000 |
| Bi-2 | | | | 0.000 | 0.000 | 0.000 | | | 0.000 | 0.0 | | |

**Figure 2.5** Dynamic Light Scattering (DLS) results for purified Hakai (aa 106-206). The apparent molecular weight of Hakai (aa 106-206) as indicated in the DLS results (~24 kDa) is twice compared the monomeric molecular weight (~ 12kDa), suggesting that Hakai (aa 106-206) exists as a dimer in solution

### 2.3.3 Crystallization and X-ray diffraction data of Hakai (aa 106-206)

The Hakai sequence spanning amino-acid residues 106–206 (aa 106–206) that contained both the RING domain and the minimum pTyr-binding domain, as represented schematically in Figure 2.8A, was purified as described and used for crystallization screening. Initial crystallization conditions were identified from the Hampton Crystal Screen I (Hampton Research, CA) and Wizard III crystal screen (Emerald Biosciences) using hanging drop vapour diffusion method at 298K. The initial crystals were very tiny and did not diffract X-rays.

A series of optimization steps were employed to improve the size and quality of the protein crystals. These included the variation of protein concentration, precipitant and salt concentration, use of additive screen, introduction of a paraffin oil barrier (Chayen, 1997) over the reservoir solution to slow down the equilibration rate to produce larger and better crystals, use of sodium thiosulphate ($Na_2S_2O_3$) to prevent intermolecular disulfide bridges (Bisacchi et al., 2006; Bordo' et al., 2001), modulation of temperature, and so on. Finally, after extensive optimization of the crystallization condition, diffracting diamond shaped crystals appeared after 10 days, and grew over a period of 20-25 days at 289K. The step-wise improvement in the crystal shape and quality through the optimization process is indicated in Figure 2.6.
The final diffraction quality crystals were obtained by mixing 2 µl of protein (30 mg/ml) with 2 µl of reservoir condition consisting of 140mM $Li_2SO_4$; 100mM Tris, pH 7.8; 15mM $Na_2S_2O_3$; 20–22% PEG 5000 MME; 1% isopropanol and 20–25% ethylene glycol through hanging-drop vapour-diffusion method at 289K. The final

optimized conditions were used to grow the SeMet Hakai (aa 106-206) crystals as well.

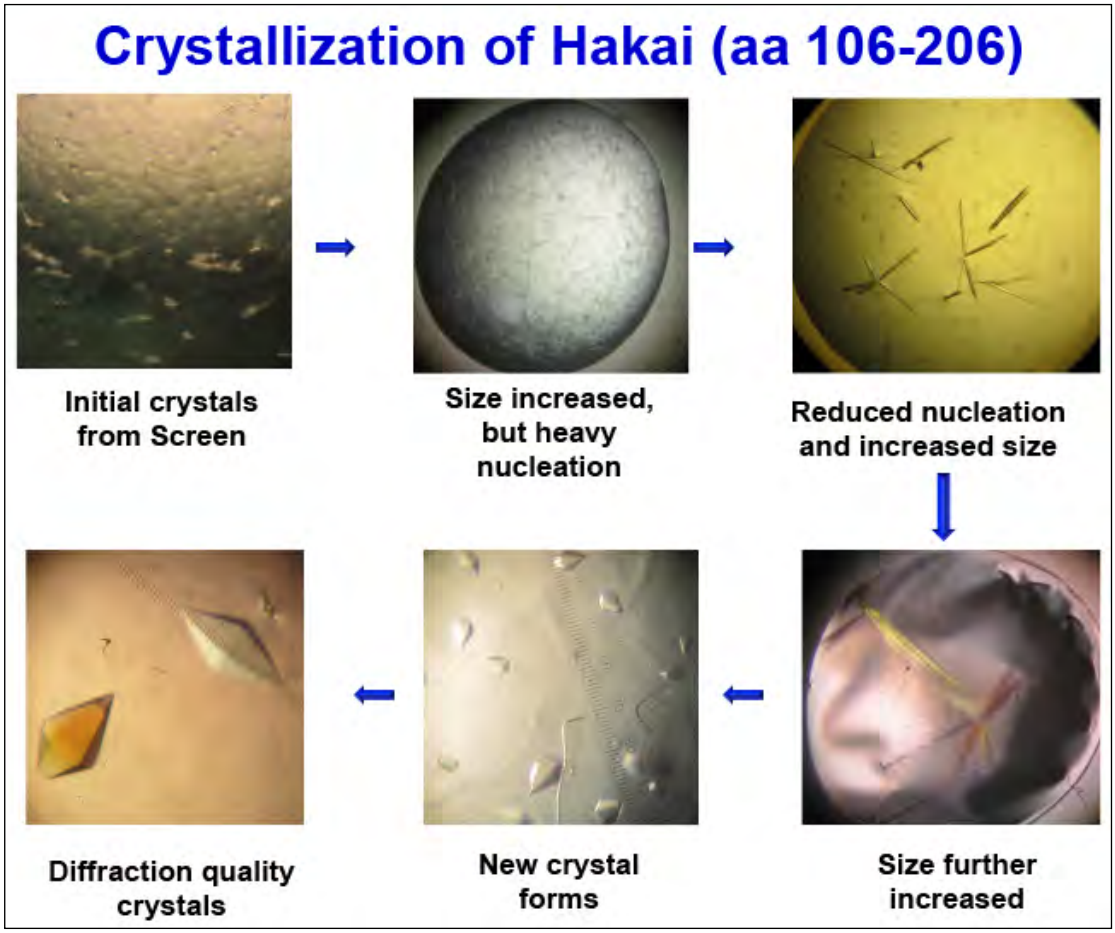



**Figure 2.6** Crystallization of Hakai (aa 106-206). Systematic and step-wise improvement in the crystal shape and quality to obtain diffraction quality of Hakai (aa 106-206)

Since the SeMet crystals of Hakai (aa 106-206) were grown in the presence of 20-25% ethylene glycol, they were directly picked from the crystallization drop and flash cooled in a $N_2$ cold stream (100 K) with no additional cryo-protection step. A complete SAD data set was collected to 1.9Å resolution at the synchrotron beamlines (Figure 2.7). The crystal belonged to $P6_222$ and the structure was solved using SAD method. The X-ray data collection and refinement statistics are provided in Table 2.2.

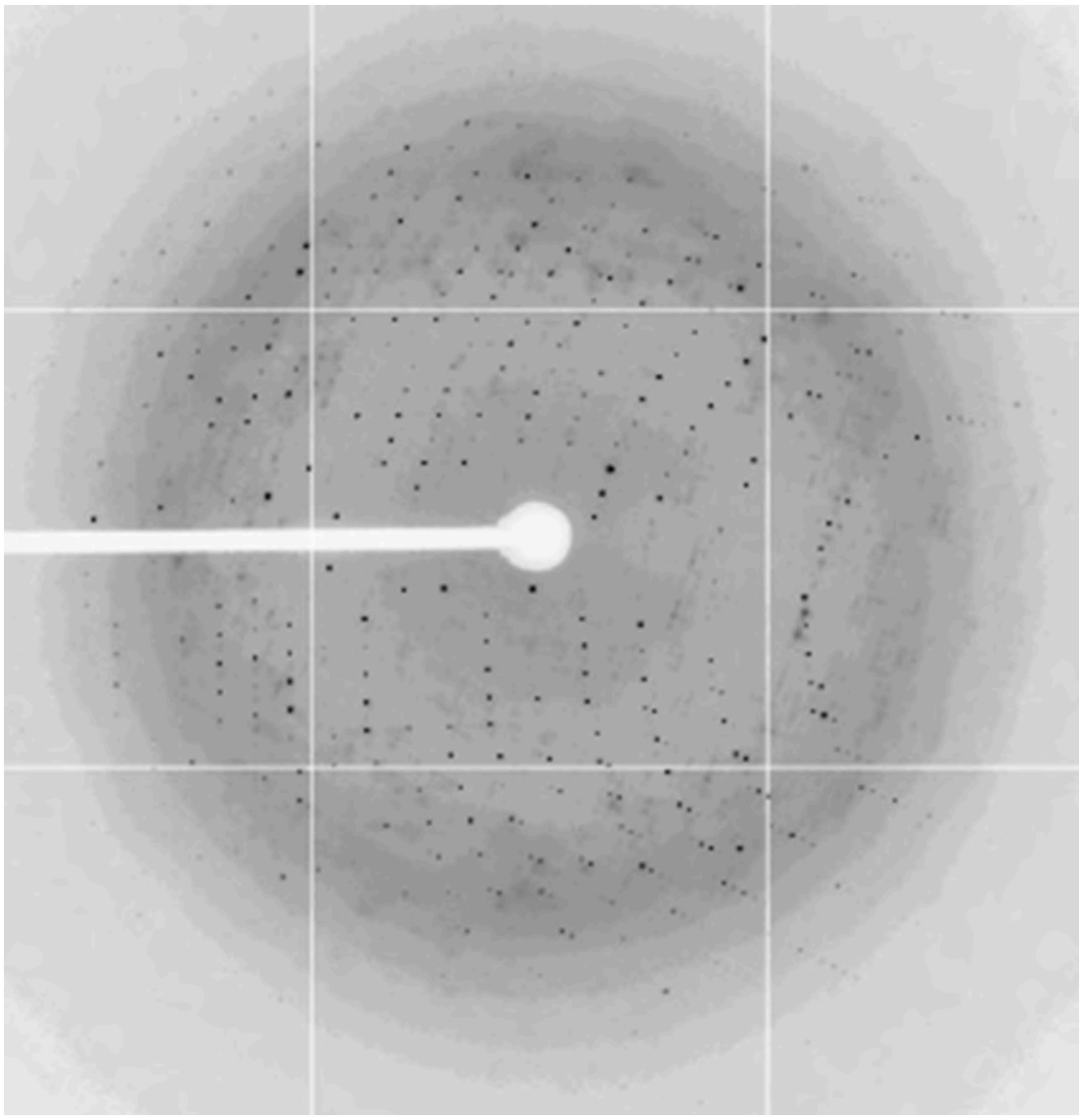


**Figure 2.7** Representative diffraction pattern of Hakai (aa 106-206) crystal. The data collected from ADSC Q315 diffractometer system at 13-B1beamline (NSRRC, Taiwan)

| | SeMet-SAD |
|---|---|
| **Data collection** | |
| Space group | $P6_222$ |
| Cell dimensions | |
| *a*, *b*, *c* (Å) | 64.66, 64.66, 121.04 |
| α, β, γ (°) | 90.00, 90.00, 120.00 |
| Wavelength (Å) | 0.979 |
| Resolution (Å) | 50-1.9 (1.97-1.90) |
| Observed reflections | 509428 |
| Unique reflections | 22371 |
| $R_{sym}$ [a] | 0.052 (0.428) |
| $I / \sigma I$ | 21.3 (12.6) |
| Completeness (%) | 99.7 (100) |
| Redundancy | 22.8 (22.6) |
| | |
| **Refinement** | |
| Resolution (Å) | 25.2-1.9 |
| Reflections (working set / test set) | 19327/2153 |
| $R_{work}$ [b] / $R_{free}$ [c] | 0.2175/0.2396 |
| No. atoms | |
| Protein | 756 |
| $Zn^{2+}$ | 3 |
| Water | 73 |
| *B*-factors (Å$^2$) | |
| Protein | 38.75 |
| $Zn^{2+}$ | 30.46 |
| Water | 44.06 |
| Ramachandran statistics | |
| Most favorable regions (%) | 90.2 |
| Additional allowed regions (%) | 9.8 |
| Generously allowed regions (%) | 0.00 |
| Disallowed regions (%) | 0.00 |
| R.m.s deviations | |
| Bond lengths (Å) | 0.008 |
| Bond angles (°) | 1.202 |

Values in parentheses are for highest-resolution shell.

[a] $R_{sym} = \Sigma|Ii - \langle I \rangle| / \Sigma|Ii|$, where $I_i$ is the intensity of the i-th measurement, and <I> is the mean intensity for that reflection.

[b] $R_{work} = \Sigma||F_{obs}|-|F_{calc}|| / \Sigma|F_{obs}|$, where $F_{calc}$ and $F_{obs}$ are the calculated and observed structure factor amplitudes, respectively.

[c] $R_{free}$ = as for $R_{work}$, but was calculated using 10% of data excluded from refinement.

**Table 2.2** Crystallographic data and refinement statistics of Hakai (aa 106-206)

### 2.3.4 Crystal structure of Hakai (aa 106-206)

#### 2.3.4.1 Overall structure

The crystal structure of Hakai (aa 106–206) was solved at 1.9Å resolution (Figure 2.8). The striking feature of the crystal structure was the formation of a dimer from paired, anti-parallel Hakai (aa 106–206) monomers. Each monomer consisted of an N-terminal RING domain, followed by the C-terminal atypical zinc-binding domain that is contained within the experimentally derived minimum pTyr-binding domain. Furthermore, each monomer contained three zinc ions at three distinct sites. Two of the three-zinc coordinating sites lie in the RING domain, while the third zinc ion coordinated with the atypical zinc-binding domains of both monomers (Figure 2.8).

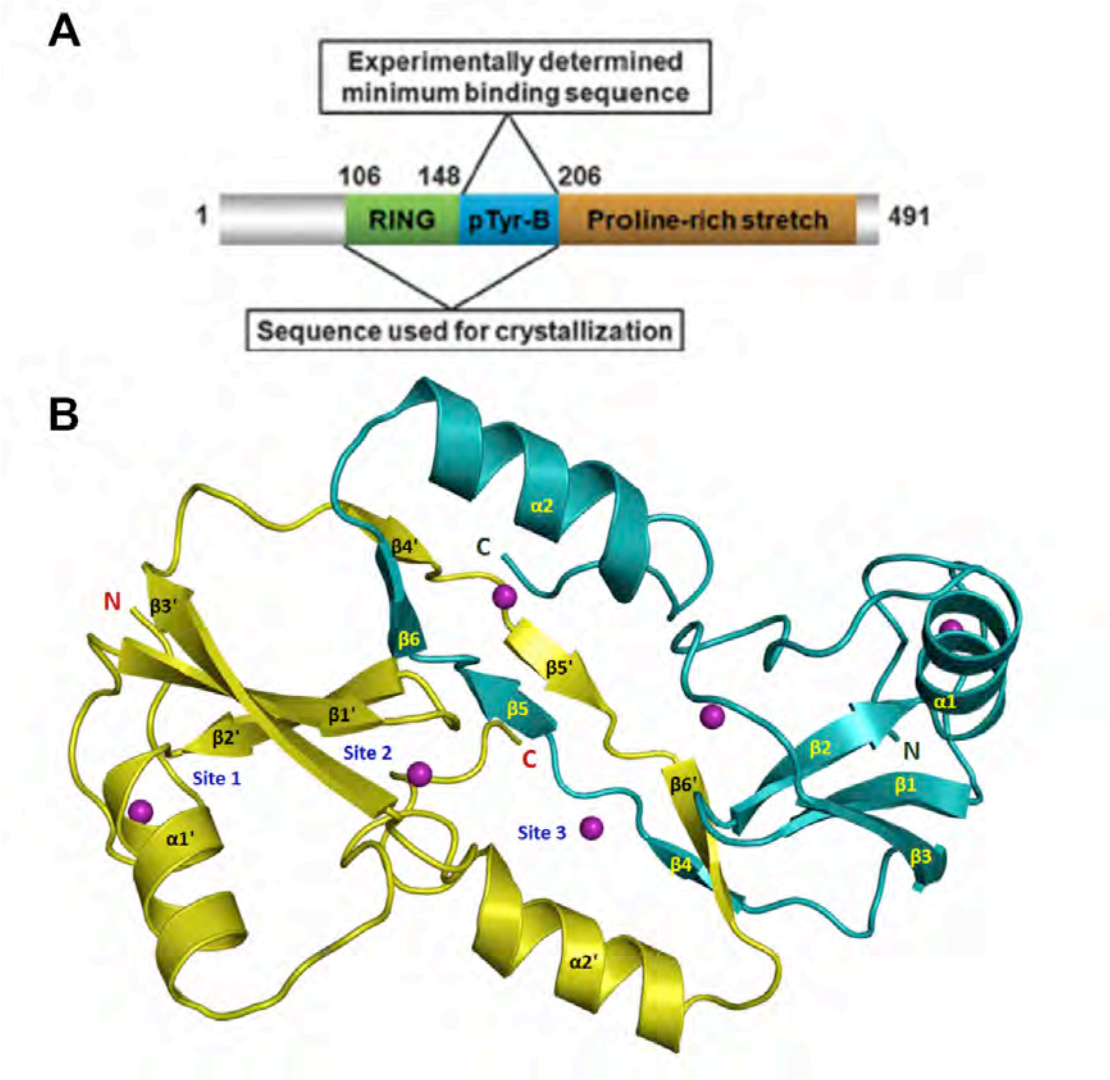



**Figure 2.8** A novel protein fold in Hakai. (A) A schematic diagram of the Hakai protein. (B) The crystal structure of Hakai (aa 106–206) reveals a dimer in an anti-parallel configuration. Each monomer contains three zinc coordination sites. Sites 1 and 2 lie in the RING domain. Site 3 is shared between the two monomers.

### 2.3.4.2 The RING domain of Hakai

The Hakai RING domain (residues 106–148) adopted a typical RING domain fold stabilized by coordinating with two zinc ions, forming a cross-brace arrangement (Figure 2.9) (Borden, 2000). The zinc- coordinating residues in the RING domain are also indicated in Figure 2.10.

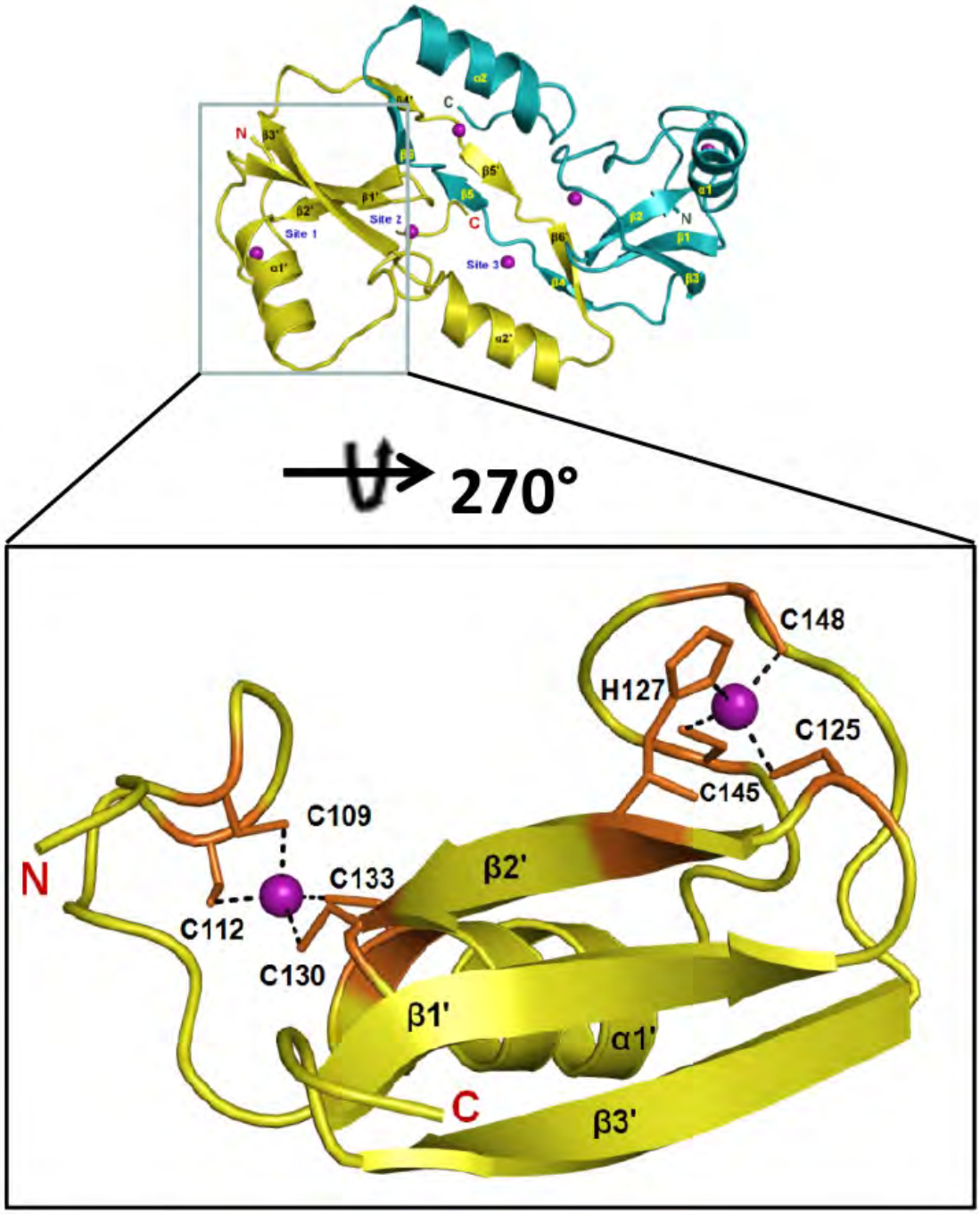



**Figure 2.9** RING domain of Hakai highlighted under the box. The coordination of zinc ions (purple spheres) by the RING domain of Hakai is shown for one of the monomers.

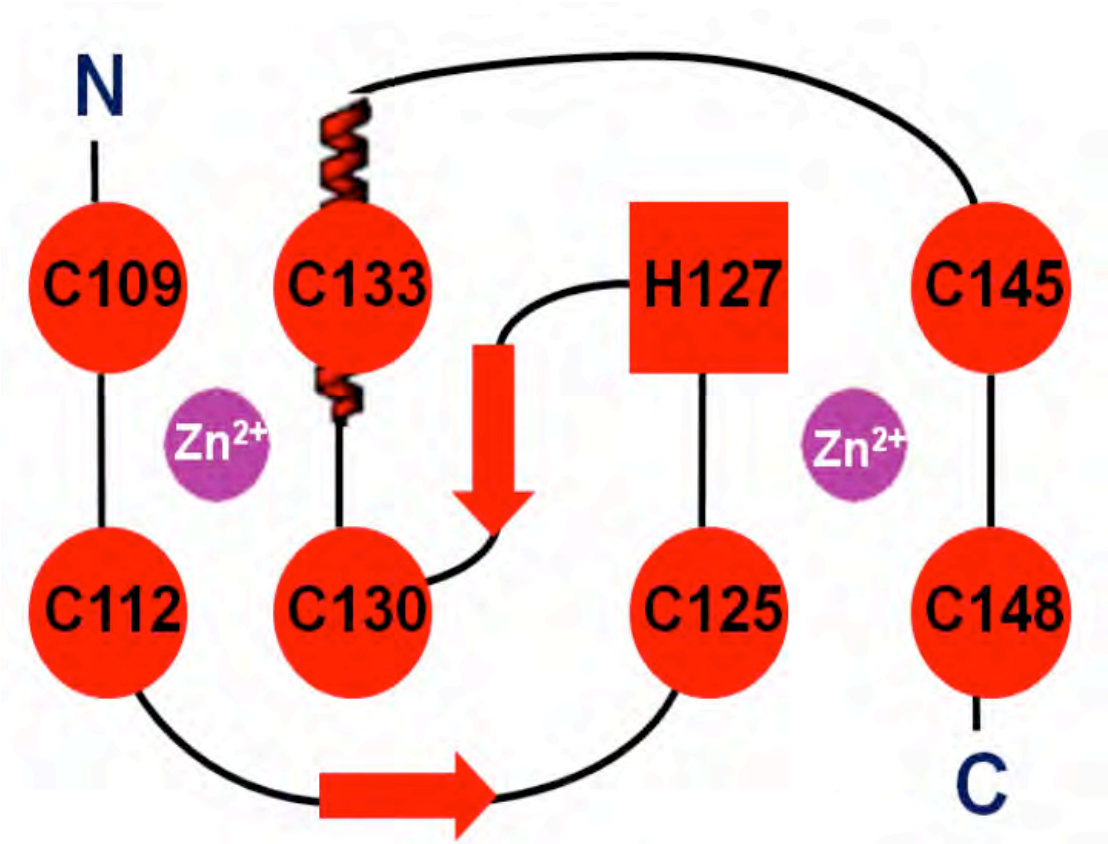



**Figure 2.10** A schematic diagram of the cross-brace arrangement of the Hakai RING domain. The Zinc ions are shown in the purple and the zinc coordinating Cysteine and Histidine residues within the RING domain are indicated.

The Hakai RING domain (aa 106-148) was compared with the existing structures in the Protein Data Bank (PDB) using the DALI server [http://ekhidna.biocenter.helsinki.fi/dali_server/] (Holm et al., 2008). The RING domain of Hakai shows the highest structural similarity to Polycomb complex protein BMI-1, PDB code 2CKL (Buchwald et al., 2006), with rmsd of 1.9Å for 42 Cα atoms (Z-score= 4.3 ; 36% sequence identity. The next significant structural similarity of the Hakai RING domain was that with the RING domain of c-Cbl, PDB code 2Y1N (Dou et al., 2012a), with rmsd of 1.8Å for 40 Cα atoms (Z-score= 4.3 ; 25% sequence identity. Interestingly, the Hakai ring domain also bears significant structural similarity with CHIP U-box domain, PDB code 2OXQ (Xu et al., 2008), with rmsd of 2.0 Å for 41 Cα atoms (Z-score= 4.3 ), although the sequence similarity is just 5%.

Thus, the analysis suggested that although the RING domain of Hakai is structurally similar to a number of the RING domains as well as U-box domain, but bears very low sequence similarity with other RING domains. Table 2.3 summarizes the DALI results indicating the representative structures of the RING and U-box domains that bear structural similarity with the Hakai RING (aa 106-148).

| DALI matches | PDB code | Z score | RMSD [Å] | Aligned Length | Sequence Identity [%] | Protein Name |
|---|---|---|---|---|---|---|
| 1 | 2CKL | 4.3 | 1.9 | 42 | 36 | Polycomb complex protein BMI-1 |
| 2 | 2Y1N | 4.3 | 1.8 | 40 | 25 | c-Cbl |
| 3 | 2QJ0 | 3.9 | 2.1 | 41 | 12 | RNF8 |
| 4 | 2XEU | 3.7 | 2.1 | 40 | 30 | Ring finger protein 4 |
| 5 | 2H0D | 3.7 | 1.9 | 41 | 32 | RING1B |
| 6 | 2OXQ | 3.6 | 2.0 | 41 | 5 | CHIP U-box complex |
| 7 | 1T1H | 3.5 | 1.9 | 41 | 15 | Armadillo repeat containing protein |
| 8 | 2VJE | 3.5 | 1.8 | 41 | 27 | MDM2 |
| 9 | 4F52 | 3.5 | 2.1 | 40 | 25 | CULLIN-1 |
| 10 | 1UR6 | 3.4 | 2.4 | 40 | 23 | CNOT4 |
| 11 | 2F42 | 3.5 | 2.1 | 41 | 5 | STIP1 Homology and U-box containing protein 1 |
| 12 | 2ECL | 2.0 | 2.5 | 39 | 31 | RING-BOX PROTEIN 2 |

**Table 2.3** Structural comparison of Hakai RING domain with structural homologs. Representative structural matches of the Hakai RING domain (aa 106-148) searched in the PDB using DALI server

### 2.3.4.3 A novel protein fold in Hakai: The minimum pTyr-binding region

The uniqueness of the Hakai (aa 106–206) region was revealed when the structure was compared with other proteins in the PDB (Protein Data Bank) using the DALI

server [http://ekhidna.biocenter.helsinki.fi/dali_server/] (Holm et al., 2008). The results show that only the RING domain of Hakai is structurally similar to RING domains of other proteins. There is, however, no similarity beyond amino-acid residue 159 of the Hakai minimum pTyr-binding domain, which is located on the dimerization interface (Figure 2.11). The minimal pTyr-binding domain of Hakai adopts a novel, three-dimensional fold and contains three β-strands (β4, β5 and β6) and a C-terminal α-helix. The β-strands β4, β5 and β6 were in an extended configuration and formed anti-parallel β-sheets with the corresponding β-strands of the monomeric partner during homodimerization (Figure 2.12).

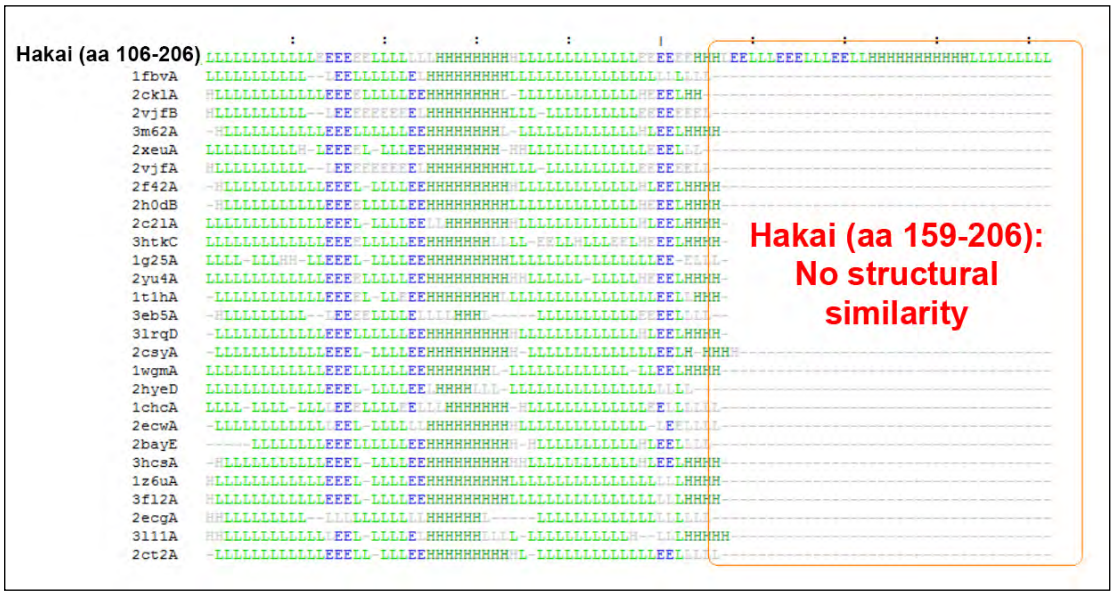



**Figure 2.11** Results of DALI search for Hakai (aa 106-206). Multiple structural alignment of Hakai (aa 106-206) with the structural matches from Protein Data Bank (PDB) using DALI server showing secondary structure assignments by DSSP (Joosten et al., 2011; Kabsch and Sander, 1983) (H: helix, E: strand, L: coil)

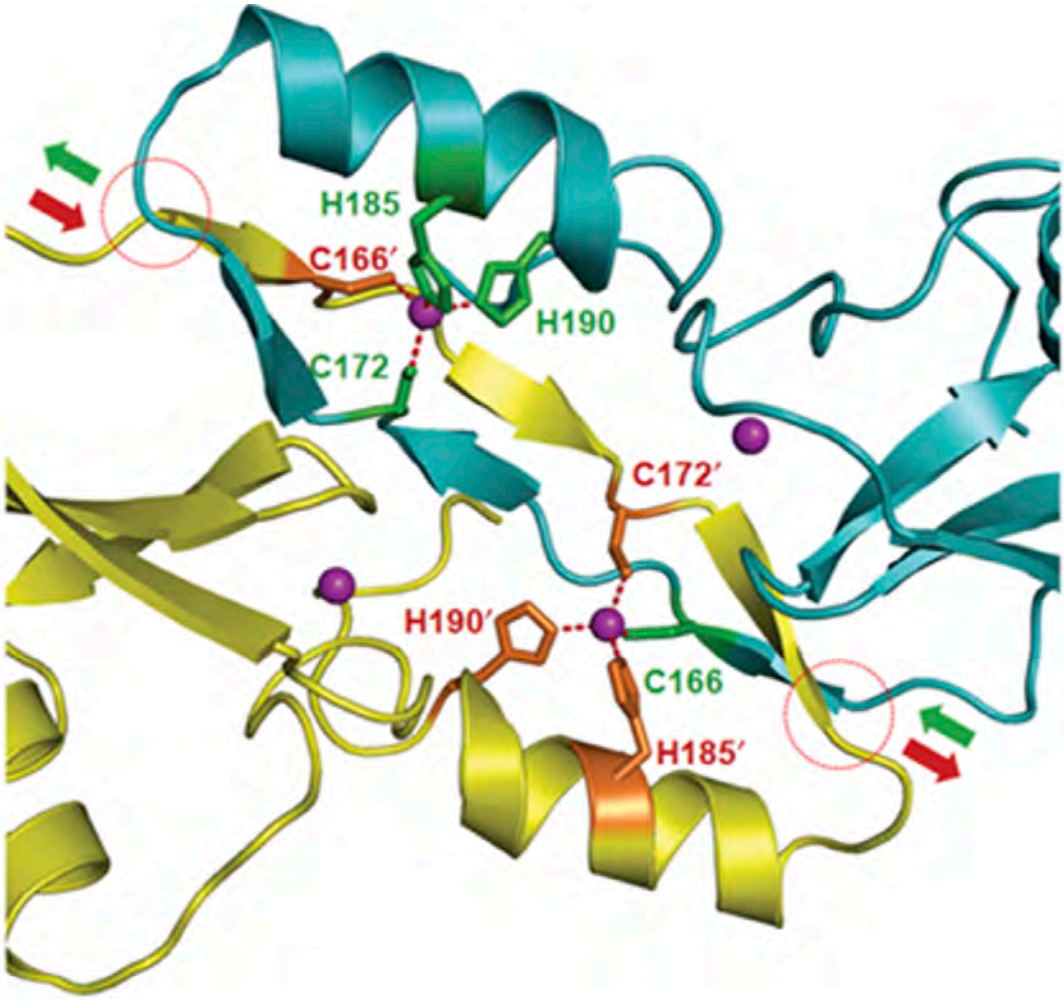



**Figure 2.12** A novel fold in the Hakai minimum pTyr-binding region. The minimum pTyr-binding domain of Hakai involves in dimerization in an intertwined configuration spanning the points indicated in circles, with the entry and exit paths shown in green and brown arrows. The zinc-interacting side chains are shown as green and brown sticks.

### 2.3.4.4 Atypical zinc-finger motif in the minimum pTyr-binding region mediates dimerization

The primary sequence of Hakai (aa106–206) suggests that each monomeric unit fulfills the required criteria of the zinc coordination consensus pattern of cysteine and histidine residues [C-x(5)-C-x(12)-H-x(4)-H] (Figure 2.13A). However, the crystal structure reveals that the single monomer unit is not capable of forming the zinc coordination sphere by itself, since a second cysteine residue is required of its anti-

parallel monomeric partner (Figure 2.13B). Thus, the structure confirms the formation of two atypical zinc-finger motifs within this region formed by two histidine residues (H185 and H190) and one cysteine residue (C172) from one monomer and a second cysteine residue (C166) from the adjacent monomer (Figure 2.13B and Figure 2.14). Therefore, each dimer contains two atypical zinc-finger motifs.

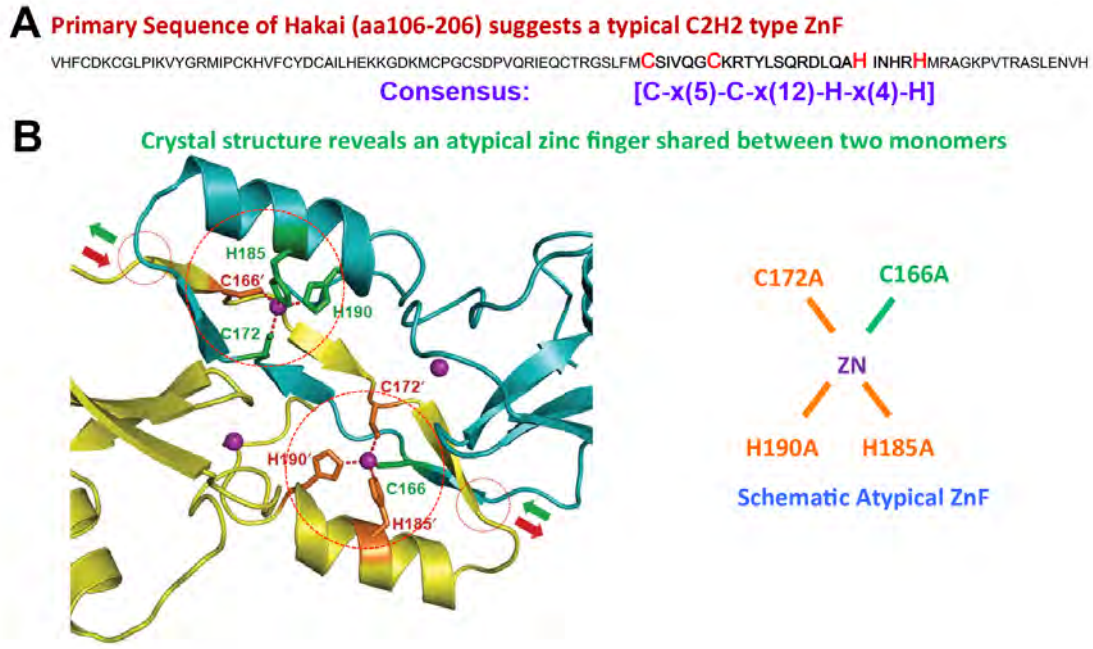



**Figure 2.13** Atypical Zinc finger motif (ZnF) in dimer interface. (A) Primary sequence of Hakai (aa 106-206) suggests the presence of the consensus sequence of a typical zinc finger. (B) Crystal structure reveals the presence of an atypical ZnF, in which two histidines and one cysteine is contributed from one monomer, while the second cysteine comes from the adjacent monomeric partner, resulting the formation of an antiparallel dimeric arrangement between the paired monomeric units

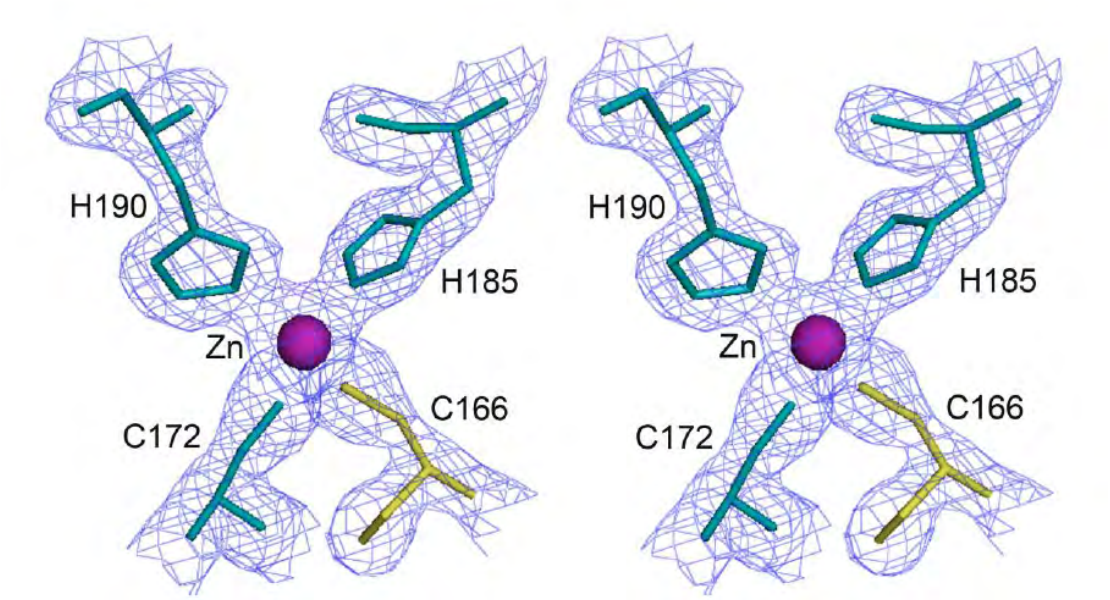



**Figure 2.14** Stereo view of the atypical zinc finger motif. The final $2F_o$-$F_c$ electron density map contoured at 1.5σ is shown in stereo view. The zinc-coordinating residues and the zinc ion at the dimer interface are shown.

### 2.3.4.5 Extensive H-bonding interactions and metal coordination defines a novel dimer interface

The crystal structure of Hakai also shows that the two Hakai monomers configure themselves to form a wide range of favorable intermolecular interactions, resulting in the formation of three β-sheets based on 12 main-chain hydrogen bonds (Figure 2.15 and Figure 2.16). This novel interlinked configuration, extensive H-bonding networks and the two atypical zinc ion interactions at the dimer interface are unique features of this distinctive homodimeric assembly.

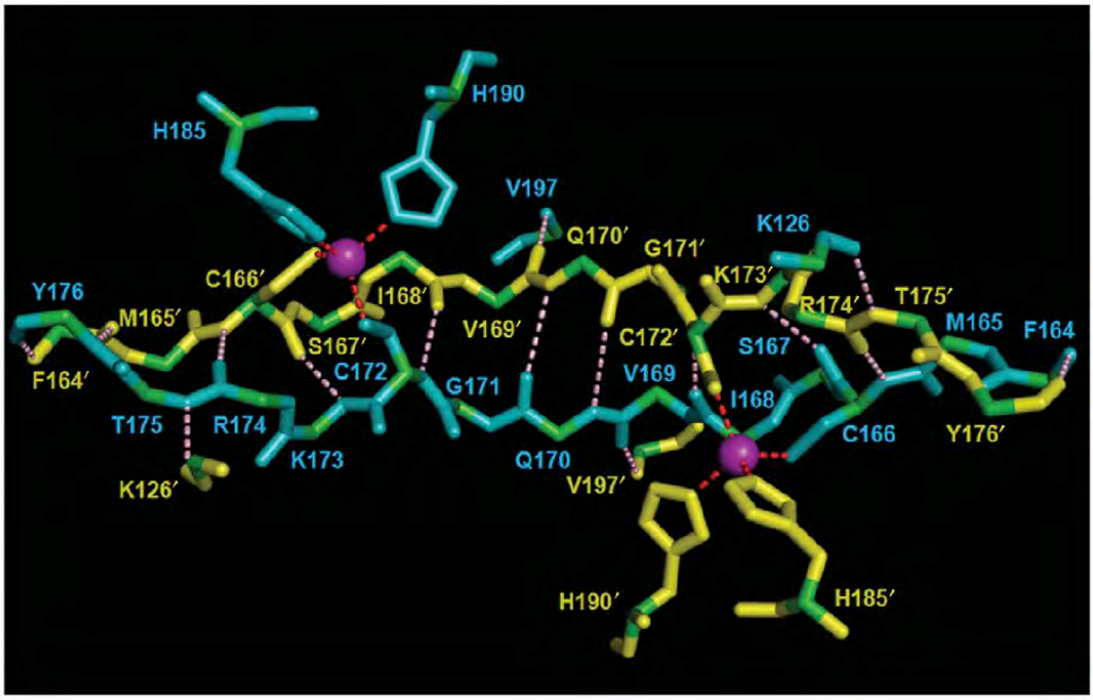



**Figure 2.15** Interactions at the dimer interface. The backbone of the Hakai (aa 106–206) residues involved in intermolecular main-chain H-bonding and the zinc-coordinating side chains of adjacent monomers at the dimer interface are shown in cyan and yellow. The pink dots indicate the main-chain H-bonds; the red dots indicate the zinc coordination bonds

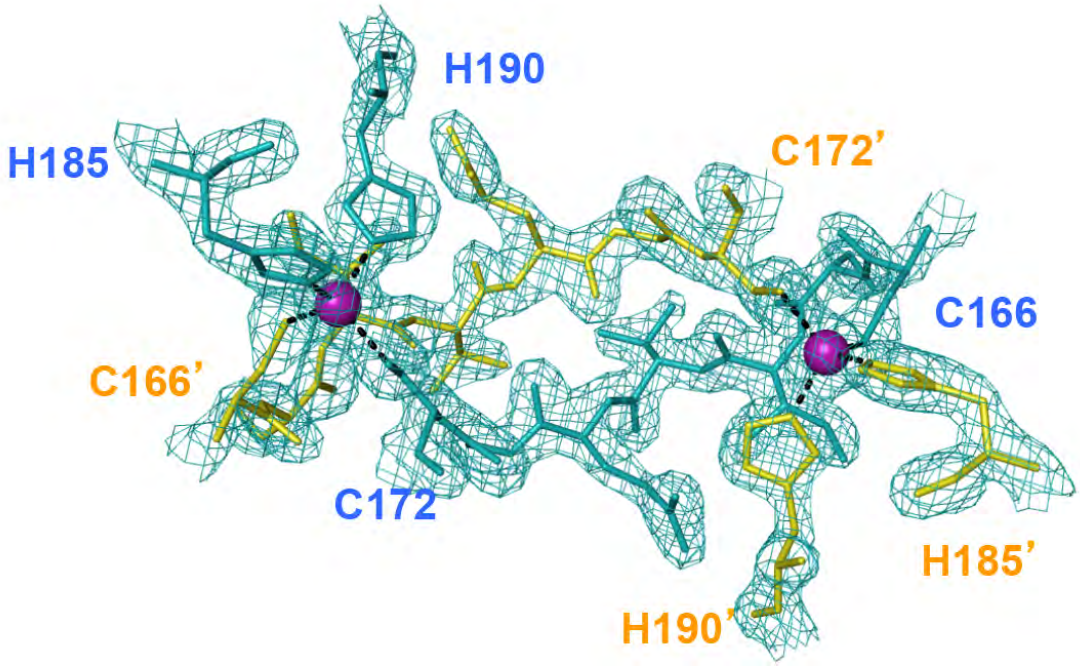



**Figure 2.16** The $2F_0$-$F_c$ map contoured at 1.5σ showing electron density of zinc coordination residues at the dimer interface of Hakai (aa 106-206).

### 2.3.4.6 Hakai dimerization involves novel intertwining of the two monomeric units

The crystal structure of Hakai also shows that the two Hakai monomers intertwine through a stretch of residues ranging from F164 to Y176 during dimerization (Figure 2.17 and Figure 2.18). A surface area of ~1650Å$^2$ of each monomer (or ~21% of each monomer surface) was formed at the dimer interface of Hakai (aa 106–206), with 34 hydrogen bond contacts between the monomers, as analyzed by the PISA server (Krissinel and Henrick, 2007).

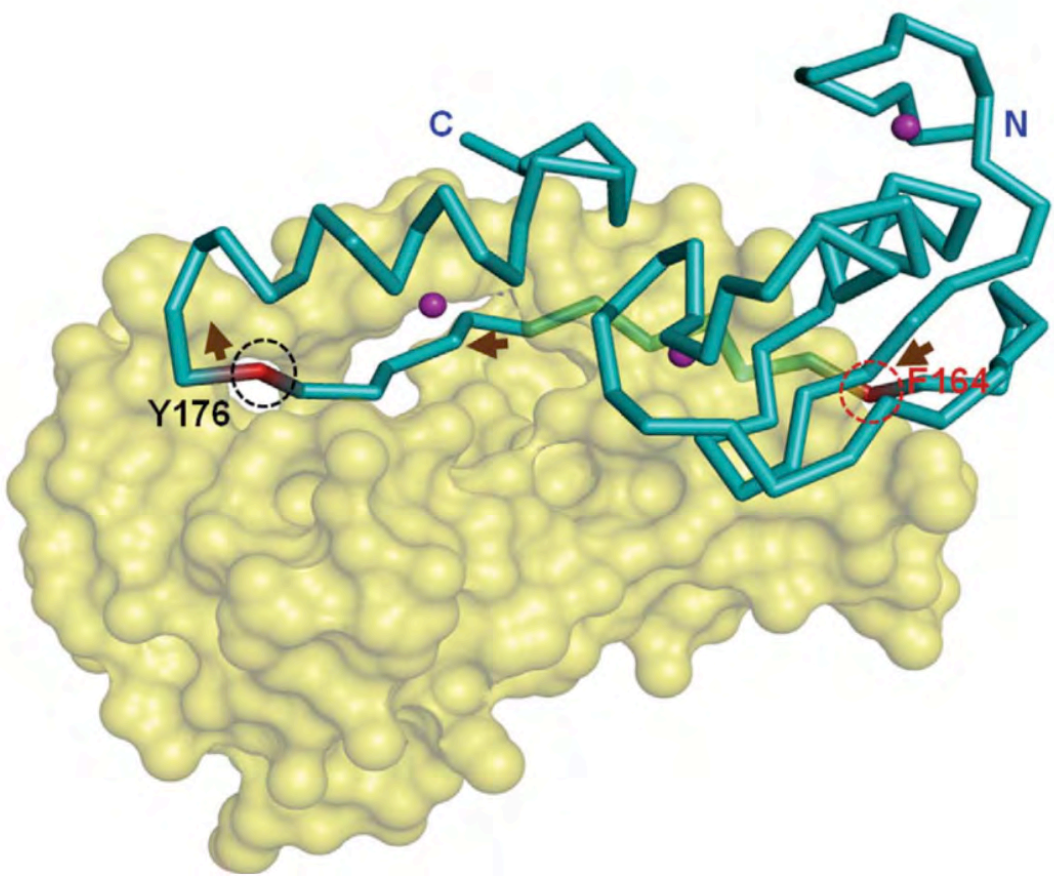



**Figure 2.17** The monomers of the interlinked Hakaidimer are shown in surface representation and Cα trace, respectively. The Cα trace monomer enters and exits the other monomer at the red and black circles, respectively. Brown arrows show its entry and exit path.

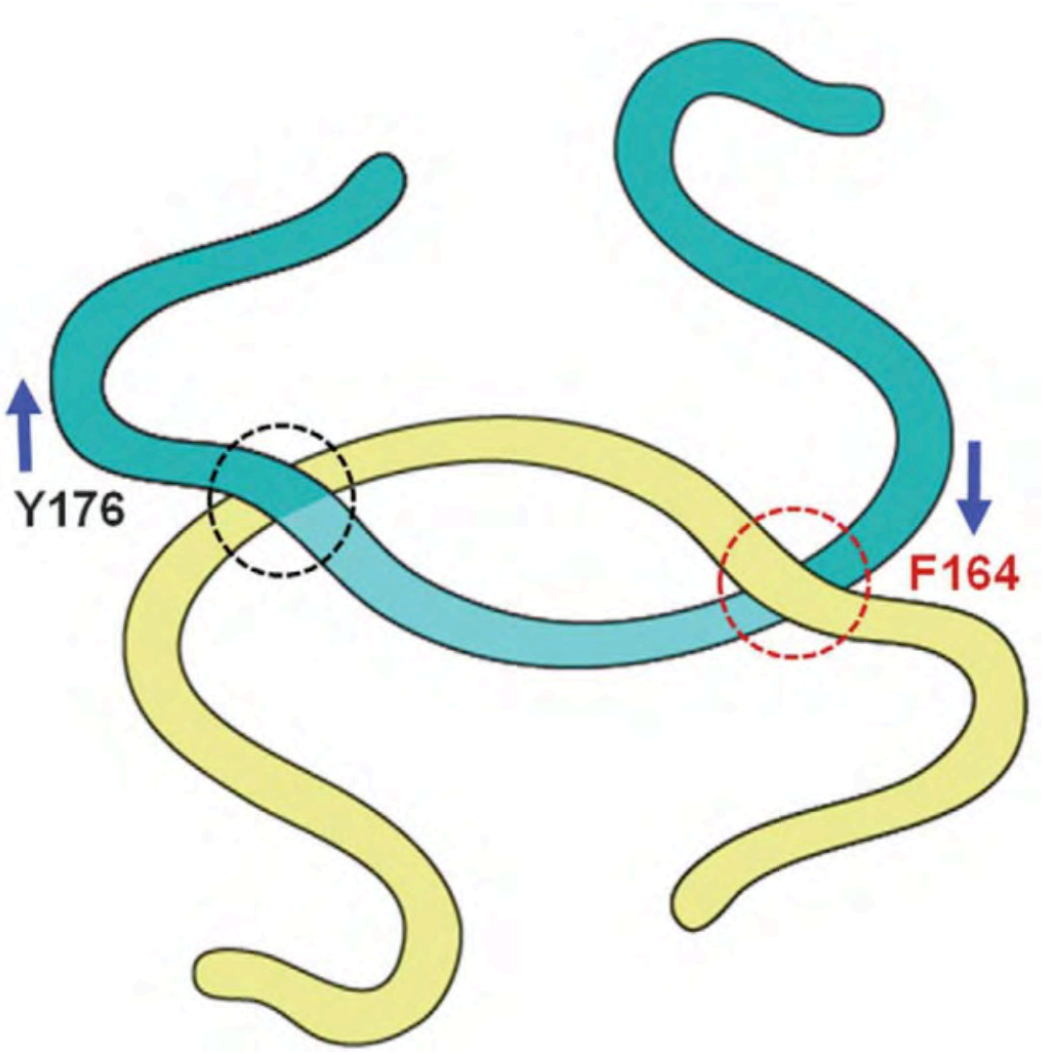



**Figure 2.18** A schematic diagram of the novel Hakai intertwined dimeric arrangement

### 2.3.4.7 Comparison with dimeric RING E3 ligases: Hakai dimer reveals a unique '*trans*' RING-RING configuration

Many members of the RING finger ubiquitin ligase family function as dimers. Dimerization generally occurs through the RING finger domain or surrounding regions and can result in homodimers or heterodimers. For heterodimers, one RING domain often lacks ligase activity and might conform and/or stabilize the active E2-binding RING domain. A list of the dimeric RING ligases whose structure has been solved is tabulated in Tables 2.4 and 2.5. A notable feature of all the known dimeric

RING ligases is the RING-RING configuration of the dimeric assembly. So far, in all the dimeric RING ligases, be it homodimer or heterodimer, the RING domains of the respective monomers face in the same side to form a '*cis*' RING-RING configuration. However, the dimeric Hakai structure revealed that the RING domains of the two monomer lie across each other in the so-called '*trans*' configuration. This type of RING configuration is likely to demonstrate novel ubiquitin conjugation and transfer mechanism from the E2 enzyme to the substrates, in comparison to other dimeric RING ligases where the RING domains lie in *cis*- configuration. A comparison of the RING-RING arrangement of Hakai with the known dimeric RING E3 ligases is shown in Figure 2.19.

| S.N. | PDB code | Description | Reference |
|---|---|---|---|
| **1** | 4AUQ | BIRC7 | (Dou et al., 2012b) |
| **2** | 2Y43 | RAD18 | (Huang et al., 2011) |
| **3** | 3EB5 | cIAP2 | (Mace et al., 2008) |
| **4** | 2YHN | IDOL | (Zhang et al., 2011) |
| **5** | 3NG2 | RNF4 | (Liew et al., 2010) |
| **6** | 1RMD | RAG1 dimerization domain | (Bellon et al., 1997) |
| **7** | 3HCS | TRAF6 | (Yin et al., 2009) |

**Table 2.4** List of homodimeric RING ligases whose structures have been solved

| S.N | PDB code | Description | Reference |
|---|---|---|---|
| **1** | 2CKL | RING1B-BMI1 heterodimer | (Buchwald et al., 2006) |
| **2** | 1JM7 | BRCA1/BARD1 heterodimer | (Brzovic et al., 2001) |
| **3** | 2VJF | Mdm2-Mdmx heterodimer | (Linke et al., 2008) |

**Table 2.5** List of heterodimeric RING ligases whose structures have been solved. These include [(MDM2 and MDMX), (BRCA1 and BARD1), (RING1b and BMI1)] For heterodimers, one RING domain (MDMX, BARD1, BMI1) often lacks ligase activity and might conform and/or stabilize the active E2-binding RING domain.

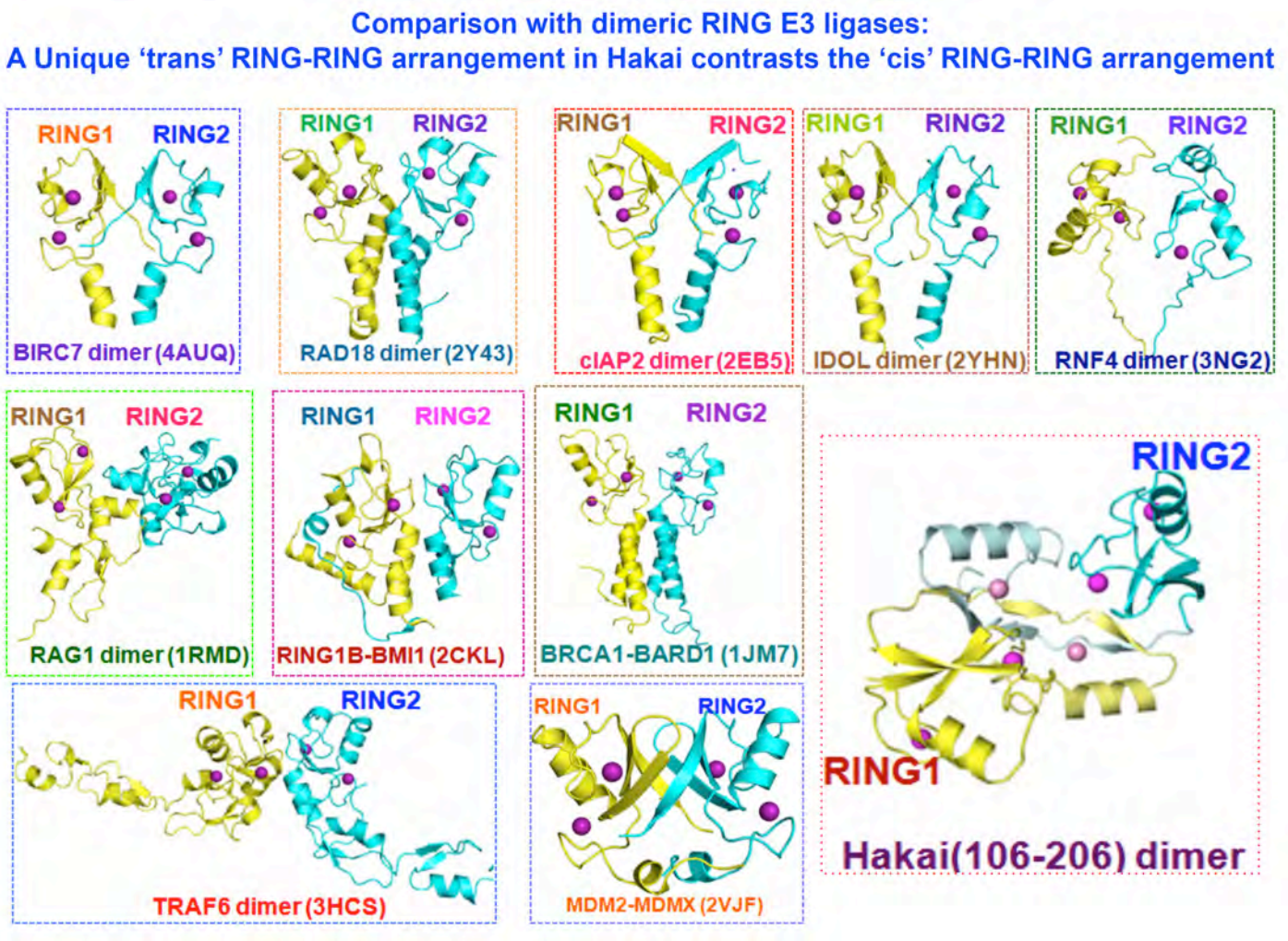



**Figure 2.19** Comparison of the RING-RING arrangement of Hakai dimer with the known dimeric RING E3 ligases. The RING domains of the respective Hakai monomers lie across each-other to form a *trans* RING-RING arrangement, whereas in all the know dimeric E3 RING ligases, the RING domains of the monomeric units face the same side to form a *cis* RING-RING configuration.

### 2.3.5 Hakai forms a dimer in solution

In addition to the crystal structure, the intermolecular NOE cross-peaks corresponding to the amides of residues at the dimer interface also indicates that Hakai (aa 106–206) forms a dimer in solution (Figure 2.20). Significantly, the structures of the dimerization interface of the RAG1 domain (Bellon et al., 1997) and Hakai (aa 106–206), as described in this study, are completely dissimilar. The formation of Hakai (aa 106–206) dimers in solution is also supported by the results obtained through dynamic light scattering, which show an apparent molecular weight of 24.1 kDa, twice than that of the monomer (Figure 2.5).

We next examined whether zinc coordination is necessary for the dimerization of Hakai and its ability to interact with its target. We first investigated whether mutations of the zinc coordinating-residues within the minimum pTyr-binding domain (C166, C172, H185 and H190) would affect the proposed Hakai dimerization. The Hakai (aa 106–206) polypeptides containing point mutations at these residues were separated on a calibrated gel-filtration column. Their gel-filtration elution profiles show that each point mutant had an apparent molecular weight equivalent to a monomeric unit of wild-type (WT) Hakai (aa 106–206) protein (~12kDa), whereas WT Hakai eluted as a dimer (Figure 2.21). Furthermore, circular dichroism performed using all the Hakai (aa 106–206) mutants ascertained that each one has maintained a well-defined secondary structure (Figure 2.22). These findings suggest that each zinc-coordinating residue is instrumental in forming the dimer interface.

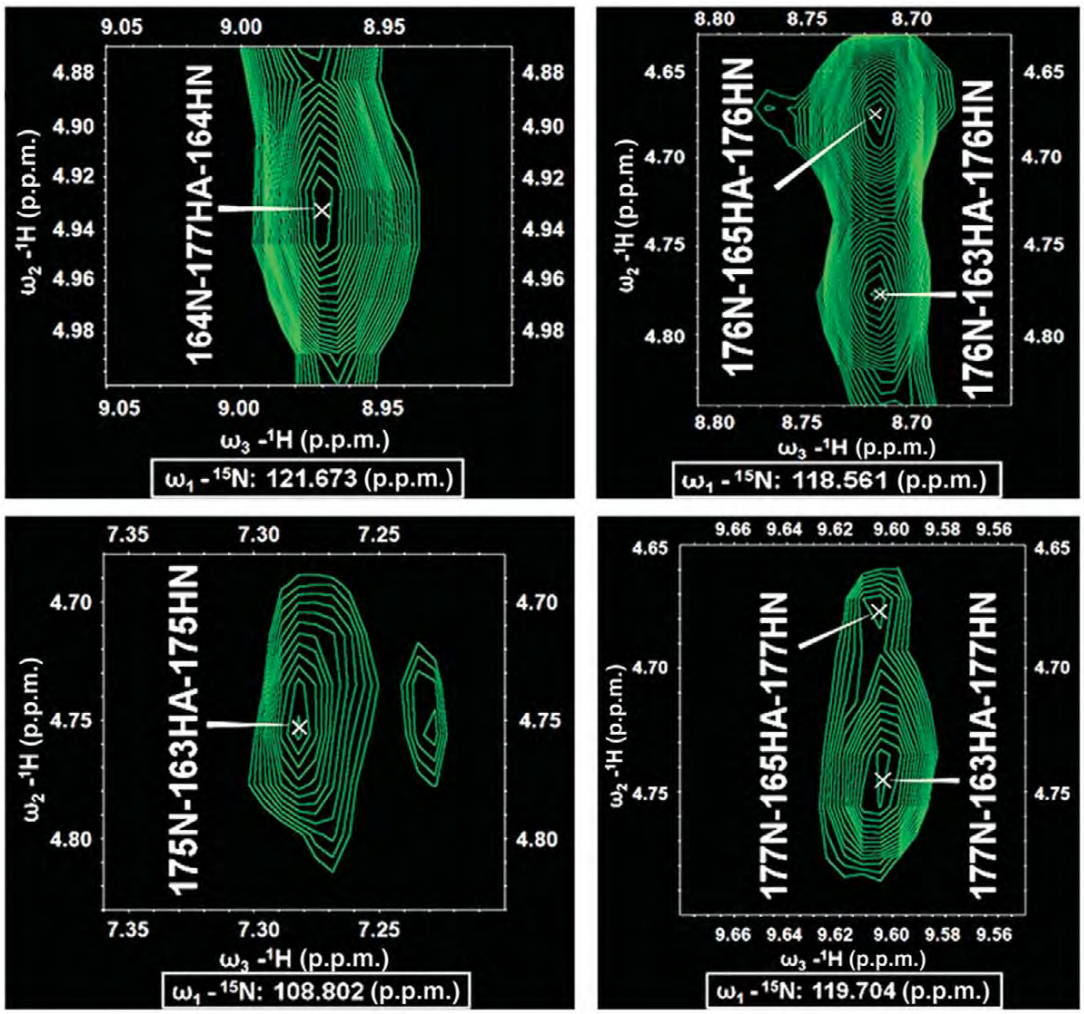



**Figure 2.20** Hakai forms a dimer in solution. A 3D $^{15}$N-NOESY spectrum showing the intermolecular NOE cross-peaks of amides corresponding to residues of Hakai (aa 106–206).

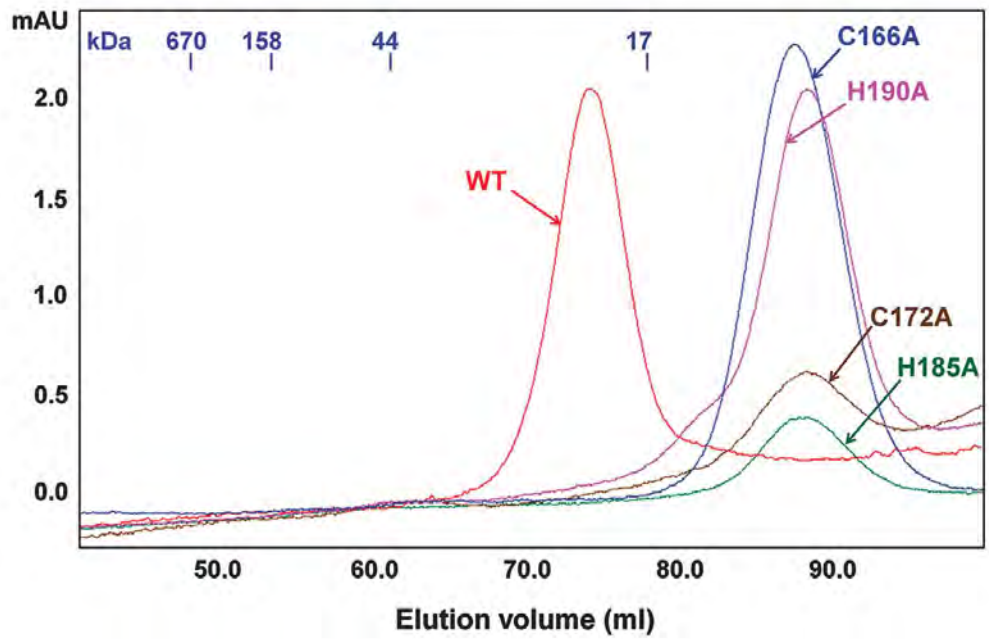



**Figure 2.21** Importance of the zinc-coordination on Hakai dimerization. WT Hakai (aa 106–206) and four Hakai (aa 106–206) point mutants were each separately used for gel-filtration chromatography. Their respective elution profiles were overlain and compared.

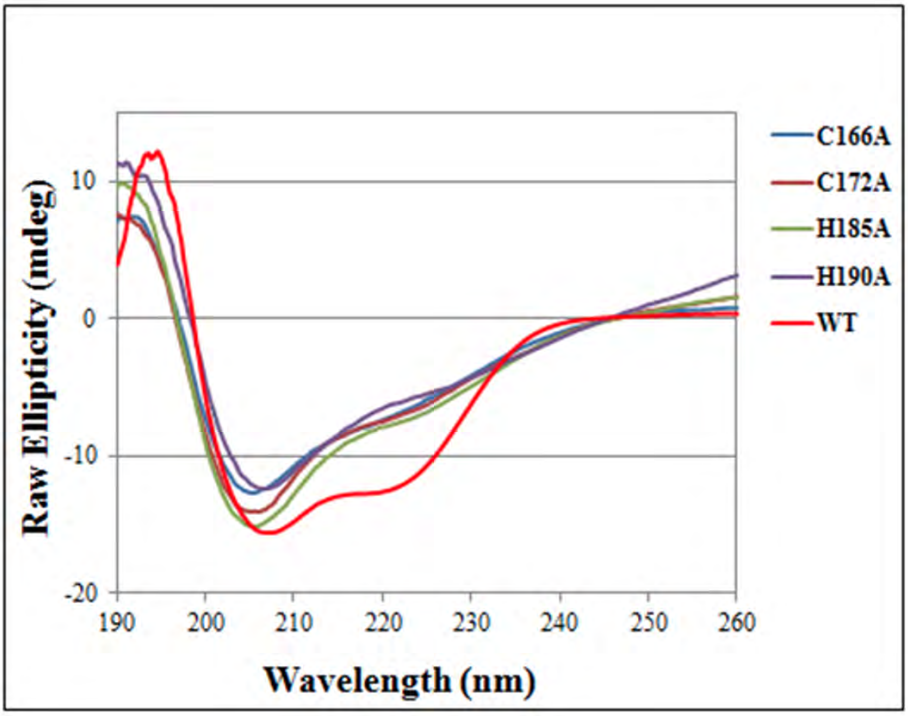



**Figure 2.22** Secondary structure of Hakai (aa 106 – 206) zinc-coordinating mutants. Circular dichroism analysis shows that each of the zinc-coordinating mutants of Hakai (aa 106 – 206) maintained a well-defined secondary structure, comparable to that of the wild type

Having determined that Hakai (aa 106–206) dimerizes in solution, we next examined if this occurs with full-length proteins. Full-length FLAG-tagged Hakai was observed to bind to its HA-tagged counterpart (Figure 2.23), indicating that dimerization also occurs between the full-length proteins.

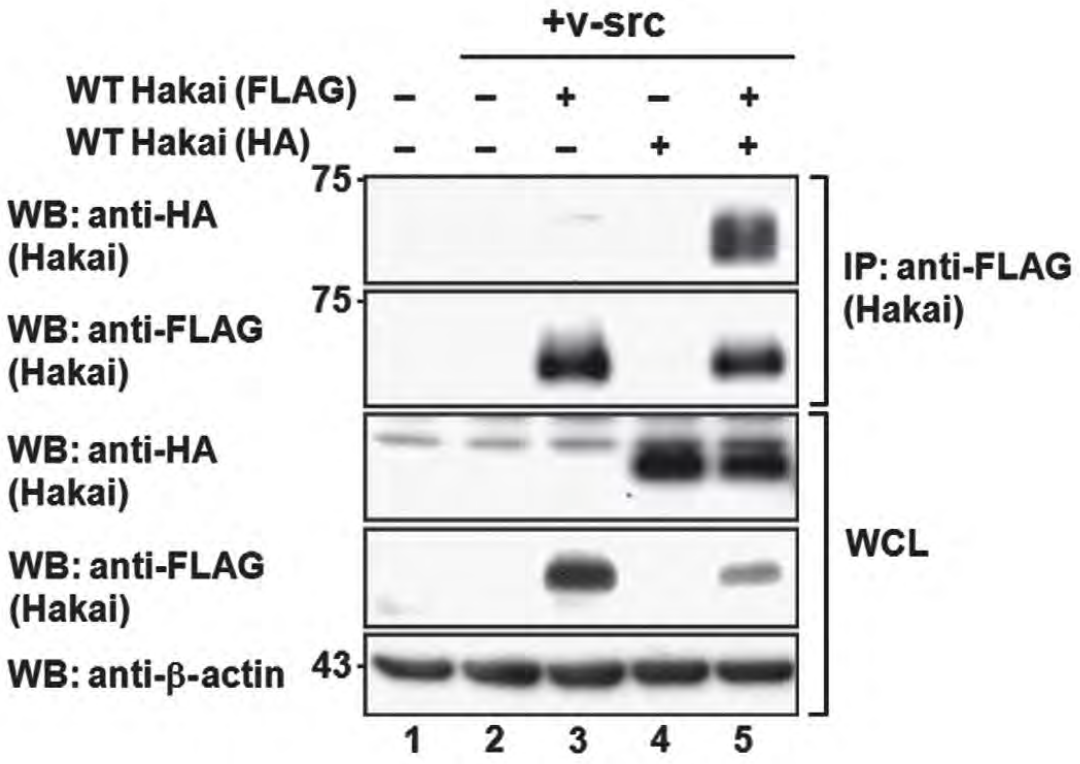



**Figure 2.23** Full-length Hakai exists as dimer inside cells. HA- and FLAG-tagged Hakai were overexpressed in the presence of Src in HEK293 cells. FLAG immunoprecipitates were analyzed for HA-tagged Hakai

To further determine whether Hakai dimerization is required for its function in binding its targets, the full-length proteins containing alanine point mutations of the zinc-coordinating residues (Figure 2.24A) were tested for their ability to bind to tyrosine-phosphorylated E-cadherin. HEK293 cells were used for such studies, as they did not express detectable levels of endogenous E-cadherin, which could have interfered with the mammalian cellular assays used. The evidence presented in Figure

2.24B indicates that none of the four Hakai point mutants interacted with tyrosine-phosphorylated E-cadherin. The collective results therefore show that zinc coordination is necessary for both dimerization of Hakai and its subsequent function in interacting with its target.

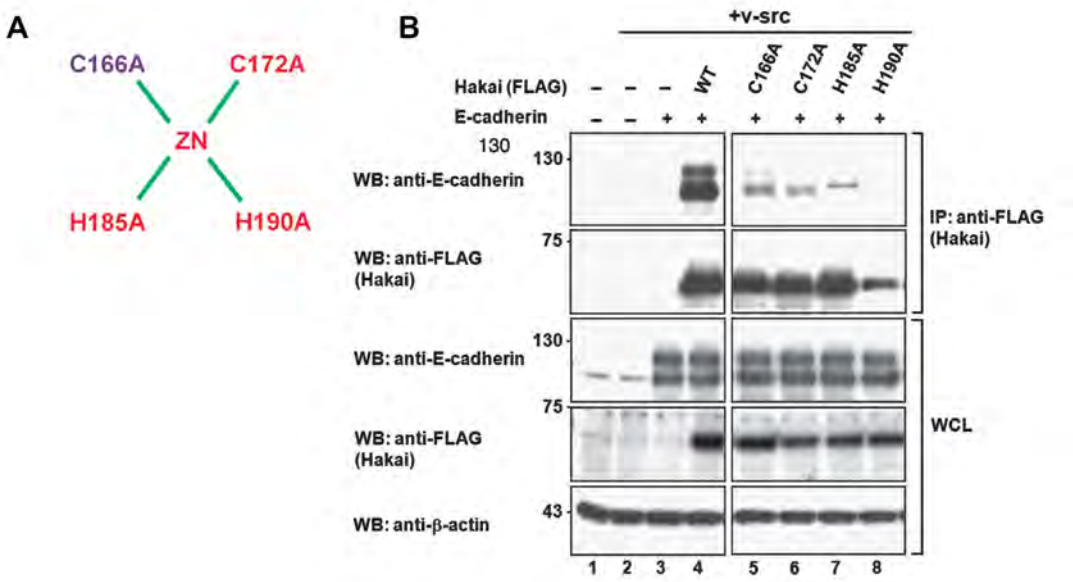



**Figure 2.24** Hakai dimerization is necessary for binding tyrosine phosphorylated E-cadherin. (A) A schematic representation of the point mutations made to C166, C172, H185 and H190 in Hakai. (B) Cell lysates containing WT Hakai or Hakai mutants were used to analyze the effects of Hakai dimerization on E-cadherin recognition.

### 2.3.6 Phosphorylation of a single tyrosine residue of the triple tyrosine-containing E-cadherin motif (YYY motif) is sufficient for binding with Hakai

Having established the novel characteristics of the Hakai zinc-coordinated homodimer, we sought to identify the target motif of this new domain. At this point, the only described target motif was in Src-phosphorylated E-cadherin. Within this motif, two (Y755 and Y756 in mouse; Y753 and Y754 in humans) of three consecutive tyrosine residues were reported to be involved in the interaction with Hakai (Fujita et al., 2002). To analyze the relative contributions of the three-tyrosine

residues in binding to Hakai, mutations were made to the tyrosine residues (Figure 2.25A). To determine which of the tyrosines were phosphorylated, we analyzed the patterns of the tyrosine phosphorylation of the point mutants after v-Src activation. The pTyr signals shown in Figure 2.25B indicate that all three adjacent tyrosines were phosphorylated.

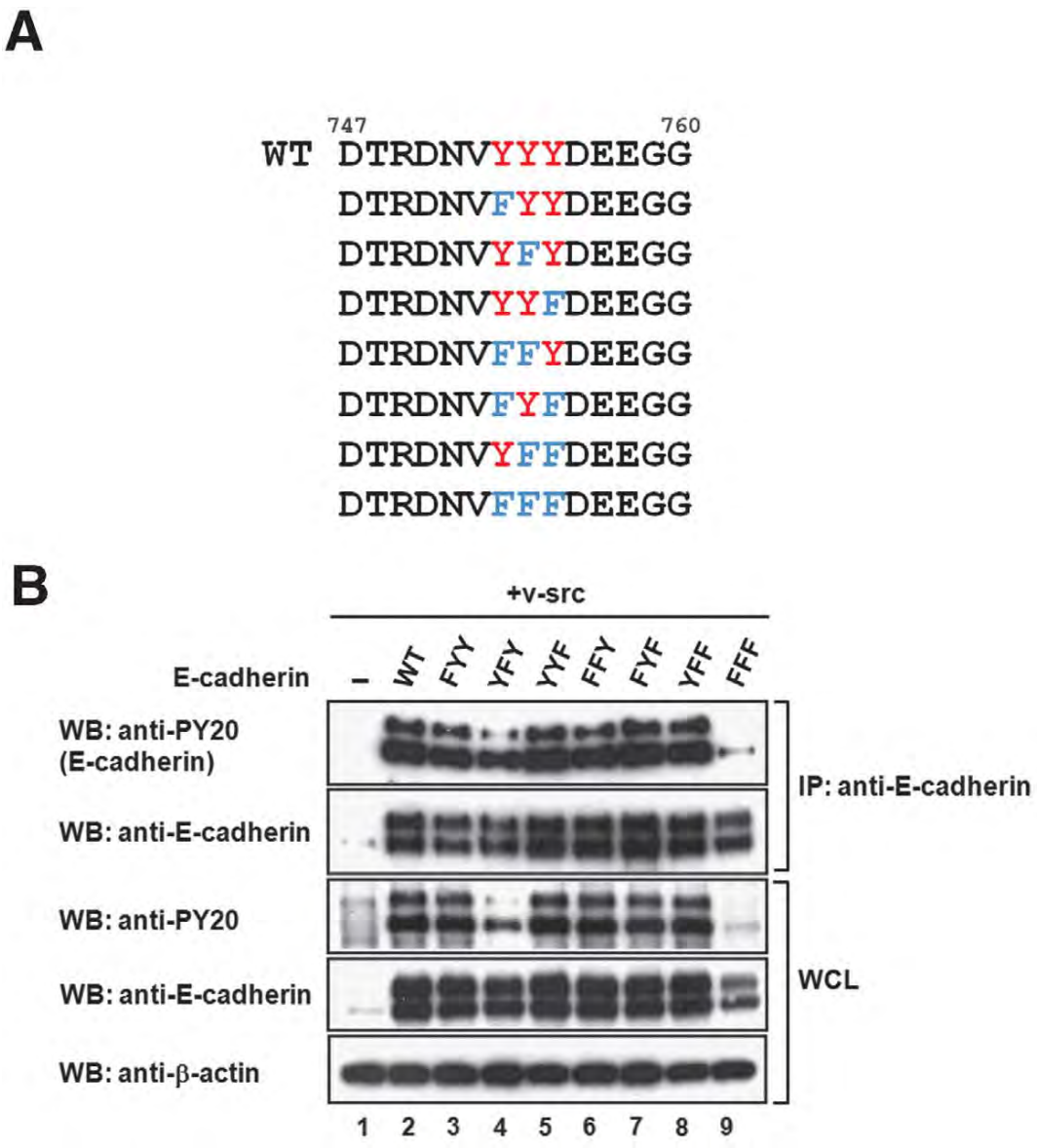
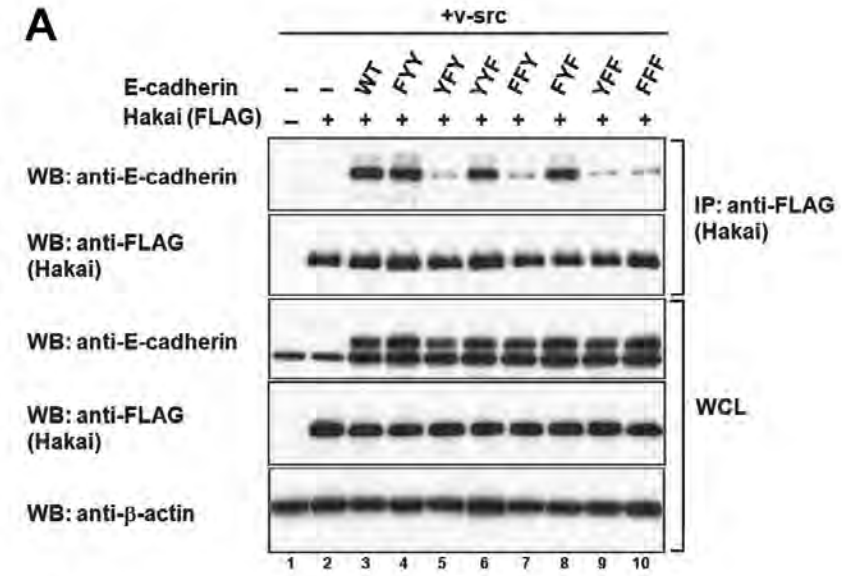
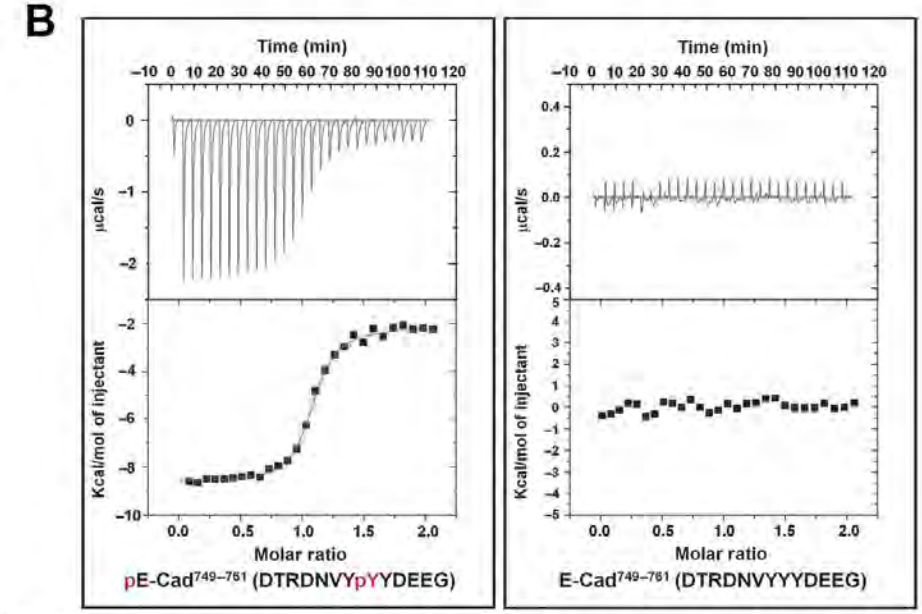



**Figure 2.25** Analysis of the pTyr signals of Src phosphorylated E-cadherin. (A) Y753, Y754 and Y755 (red) of E-cadherin were mutated to phenylalanine (blue) in different combinations. (B) The WT E-cadherin and the mutants shown in (A) were overexpressed in HEK293 cells with v-Src to analyze their pTyr signals.

We next examined the importance of the three-tyrosine residues in E-cadherin for its interaction with Hakai. The E-cadherin mutants used in the earlier experiment were analyzed for their potential to bind to WT Hakai. The results shown in Figure 2.26A indicate that Y754 is the only tyrosine significantly involved in binding; each mutant containing a substitution in this position did not bind Hakai, whereas all other mutants showed significant binding. These combined results also show that while Src binds to and phosphorylates most of the E-cadherin mutants, all the mutants with a Y754F substitution do not bind to Hakai, even when phosphorylated. This implies that the interaction between Hakai and E-cadherin depends on the direct recognition of specific pTyr residues on E-cadherin by Hakai. To verify the necessity of Y754 phosphorylation for the interaction between E-cadherin and Hakai, isothermal titration calorimetry (ITC) was performed using phosphorylated and non-phosphorylated E-cadherin peptides corresponding to aa 749–761 with Hakai (aa 106–206). The results in Figure 2.26B show that binding occurred only with the phosphorylated peptide. Furthermore, the results (2.26B) also indicate that only one E-cadherin peptide binds to the Hakai (aa 106–206) dimer at any one time.




| Peptide | Sequence | *N* | $K_a$ (x10$^6$M$^{-1}$) | $K_d$ (µM) | ΔH (kcal/mol) | TΔS (kcal/mol) | ΔG (kcal/mol) |
|---|---|---|---|---|---|---|---|
| pE-cadherin | DTRDNVYpYYDEEG | 1.008 ± 0.033 | 0.1385 ± 0.0398 | 7.22 | −12.29 ± 0.5838 | −5.18 | −7.11 |
| E-cadherin | DTRDNVYYYDEEG | No binding | | | | | |

**Figure 2.26** Phosphorylation of a single Tyrosine of E-cadherin (Y754) is critical for binding. (A) HEK293 cells were co-transfected with WT E-cadherin or its mutants together with Hakai to identify the tyrosine residues recognized by Hakai. (B) The Y754-phosphorylated and non-phosphorylated E-cadherin peptides were titrated against Hakai (aa 106–206) using ITC. The top panels show the heat release profiles after baseline correction and the lower panels indicate the binding isotherms for the interactions. The dissociation constant ($K_d$) and binding stoichiometry (N) are shown in the table.

### 2.3.7 Hakai domain recognizes acidic residues

In a similar manner, we analyzed the importance of the amino-acid residues flanking the tyrosines for the interaction between the two proteins via an alanine scan (Figure 2.27A). The immunoprecipitation results in Figure 2.27B show that there were profound contributions from aspartic acid D756 and glutamic acid E757, and significant contributions from valine V752 and aspartic acid D750. Consequently, a cluster of negative charges from the acidic amino acids is formed around the centrally binding tyrosine 754 of E-cadherin.

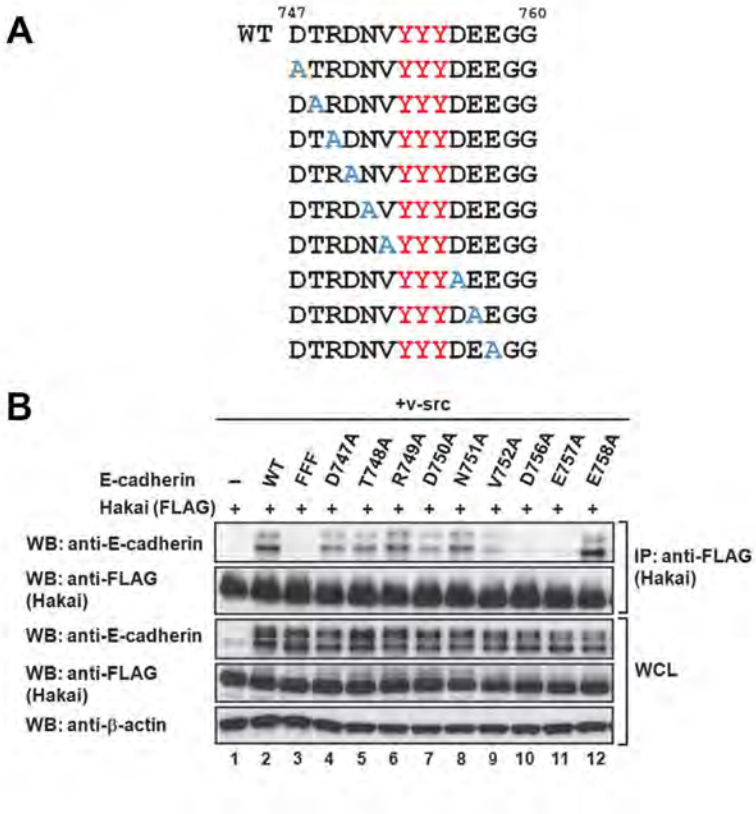



**Figure 2.27** Hakai pTyr-binding domain binds acidic residues. (A) The E-cadherin aa 747–758 were each substituted with alanine. (B) The E-cadherin mutants from (A) and Hakai were co-transfected into HEK293 cells to identify the target motif on E-cadherin. The results suggest that the central phosphotyrosine residue is most important for binding with Hakai.

### 2.3.8 Identification of novel substrates of Hakai pTyr-binding domain

To verify the motif recognized by Hakai, additional target proteins were identified. This identification was accomplished by analyzing Hakai-binding proteins phosphorylated by Src using mass spectrometry. A list of the proteins obtained from a typical experiment is appended below and Figure 2.28 provides details about the identification of cortactin as a potential target. The list

1. Protein phosphatase 1B isoform 1
2. Protein phosphatase 1A
3. Protein phosphatase 1 regulatory subunit 12A
4. casein kinase I delta
5. Casein kinase I isoform epsilon
6. **CTTN Src substrate cortactin**
7. **DOK1**
8. cAMP-dependent protein kinase type I-alpha regulatory subunit
9. ILF2 Interleukin enhancer-binding factor 2
10. IRS4 Insulin receptor substrate 4
11. IRAK1 interleukin-1 receptor-associated kinase 1 isoform 3
12. JAK1 Tyrosine-protein kinase JAK1
13. Protein kinase C iota
14. Protein kinase C delta
15. Isoform 1 of Nck-associated protein 1 mammary tumor-associated protein INT6
16. ARHGAP21 Rho GTPase activating protein 21
17. Human rab GDI
18. LIM domains containing 1
19. phosphoinositide-3-kinase, regulatory subunit 4
20. Phosphatidylinositol 4-kinase alpha
21. E3 ubiquitin-protein ligase BRE1B RNF40
22. E3 ubiquitin-protein ligase BRE1A RNF20

**Figure 2.28.** Hakai interacts with Cortactin. Mascot spectra read-out.

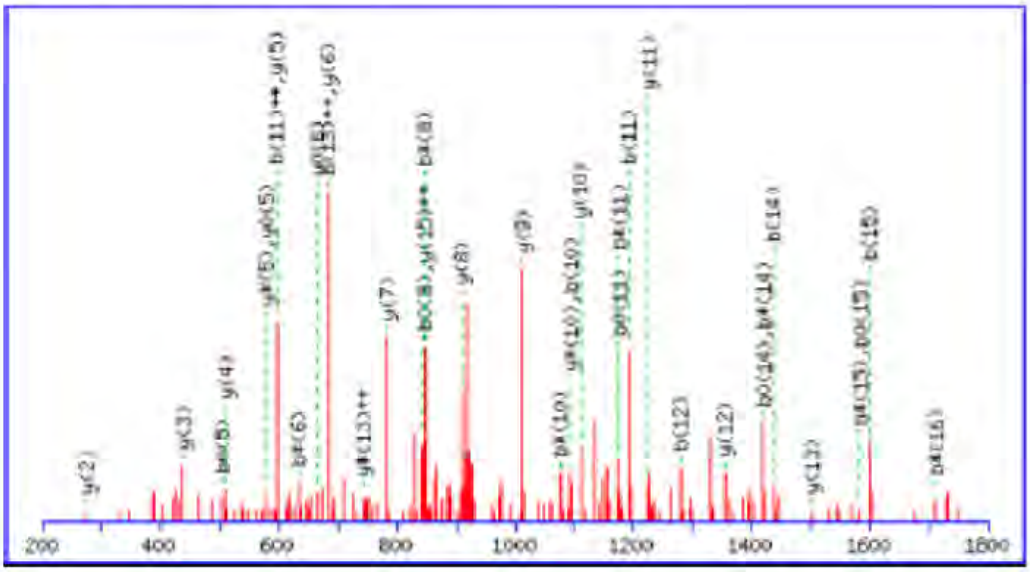
MATRIX SCIENCE **Mascot Search Results**

**Peptide View**

MS/MS Fragmentation of NASTFEDVTQVSSAYQK
Found in IPI00029601, Tax_Id=9606 Gene_Symbol=CTTN Src substrate cortactin

Match to Query 8788: 1873.880268 from(937.947410,2+)
Title: 110105_JR_Permeen_06.2199.2199.2.dta
Data file C:\TempX1\110105_JR_Permeen_CS20.Mascot

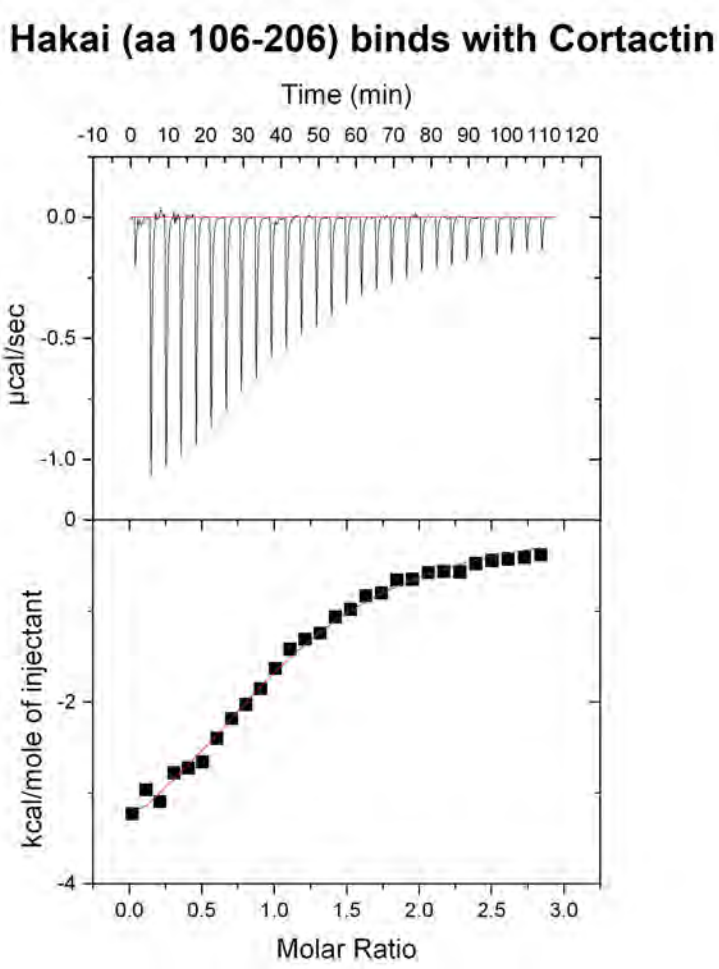
Click mouse within plot area to zoom in by factor of two about that point
Or, Plot from 200 to 1800 Da Full range




Monoisotopic mass of neutral peptide Mr(calc): 1873.8694
Fixed modifications: Carbamidomethyl (C)
Ions Score: 56 Expect: 0.00088
Matches (Bold Red): 37/180 fragment ions using 82 most intense peaks

| # | b | $b^{++}$ | $b^{*}$ | $b^{*++}$ | $b^{0}$ | $b^{0++}$ | Seq. | y | $y^{++}$ | $y^{*}$ | $y^{*++}$ | $y^{0}$ | $y^{0++}$ | # |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 115.0502 | 58.0287 | 98.0237 | 49.5155 | | | N | | | | | | | 17 |
| 2 | 186.0873 | 93.5473 | 169.0608 | 85.0340 | | | A | 1760.8337 | 880.9205 | 1743.8072 | 872.4072 | 1742.8232 | 871.9152 | 16 |
| 3 | 273.1193 | 137.0633 | 256.0928 | 128.5500 | 255.1088 | 128.0580 | S | 1689.7966 | **845.4019** | 1672.7701 | 836.8887 | 1671.7861 | 836.3967 | 15 |
| 4 | 374.1670 | 187.5872 | 357.1405 | 179.0739 | 356.1565 | 178.5819 | T | 1602.7646 | 801.8859 | 1585.7380 | 793.3727 | 1584.7540 | 792.8807 | 14 |
| 5 | 521.2354 | 261.1214 | 504.2089 | 252.6081 | **503.2249** | 252.1161 | F | **1501.7169** | 751.3621 | 1484.6904 | **742.8488** | 1483.7064 | 742.3568 | 13 |
| 6 | 650.2780 | 325.6427 | **633.2515** | 317.1294 | 632.2675 | 316.6374 | E | **1354.6485** | 677.8279 | 1337.6220 | 669.3146 | 1336.6379 | 668.8226 | 12 |
| 7 | 765.3050 | 383.1561 | 748.2784 | 374.6429 | 747.2944 | 374.1508 | D | **1225.6059** | 613.3066 | 1208.5794 | 604.7933 | 1207.5953 | 604.3013 | 11 |
| 8 | 864.3734 | 432.6903 | **847.3468** | 424.1771 | **846.3628** | 423.6850 | V | **1110.5790** | 555.7931 | **1093.5524** | 547.2798 | 1092.5684 | 546.7878 | 10 |
| 9 | 965.4211 | 483.2142 | 948.3945 | 474.7009 | 947.4105 | 474.2089 | T | **1011.5106** | 506.2589 | 994.4840 | 497.7456 | 993.5000 | 497.2536 | 9 |
| 10 | **1093.4796** | 547.2435 | **1076.4531** | 538.7302 | 1075.4691 | 538.2382 | Q | **910.4629** | 455.7351 | 893.4363 | 447.2218 | 892.4523 | 446.7298 | 8 |
| 11 | **1192.5481** | **596.7777** | **1175.5215** | 588.2644 | **1174.5375** | 587.7724 | V | **782.4043** | 391.7058 | 765.3777 | 383.1925 | 764.3937 | 382.7005 | 7 |
| 12 | **1279.5801** | 640.2937 | 1262.5535 | 631.7804 | 1261.5695 | 631.2884 | S | **683.3359** | 342.1716 | 666.3093 | 333.6583 | **665.3253** | 333.1663 | 6 |
| 13 | 1366.6121 | **683.8097** | 1349.5856 | 675.2964 | 1348.6016 | 674.8044 | S | **596.3039** | 298.6556 | **579.2773** | 290.1423 | **578.2933** | 289.6503 | 5 |
| 14 | **1437.6492** | 719.3283 | **1420.6227** | 710.8150 | **1419.6387** | 710.3230 | A | **509.2718** | 255.1396 | 492.2453 | 246.6263 | | | 4 |
| 15 | **1600.7126** | 800.8599 | **1583.6860** | 792.3466 | **1582.7020** | 791.8546 | Y | **438.2347** | 219.6210 | 421.2082 | 211.1077 | | | 3 |
| 16 | 1728.7711 | 864.8892 | **1711.7446** | 856.3759 | 1710.7606 | 855.8839 | Q | **276.1714** | 138.0893 | 258.1448 | 129.5761 | | | 2 |
| 17 | | | | | | | K | 147.1128 | 74.0600 | 130.0863 | 65.5468 | | | 1 |

Mouse cortactin is phosphorylated by Src primarily on Y482 and Y485 (Ren et al., 2009). Interestingly, these two tyrosines are also surrounded by several acidic residues. Initially, Isothermal titration calorimetry (ITC) was used to verify the binding of this Cortactin motif that contains the two phosphorylated tyrosines surrounded by acidic residues binds with Hakai pTyr-bind domain. The result shown in Figure 2.29 clearly reveals that the Hakai pTyr-binding domain binds with Cortactin with significant affinity.




| Peptide | Sequence | N | $K_a$(x$10^6$ $M^{-1}$) | $K_d$(µM) | ΔH(kcal/mol) | TΔS(kcal/mol) | ΔG(kcal/mol) |
|---|---|---|---|---|---|---|---|
| Cortactin$^{477-489}$ | AEDDTpYDGpYESDL | 0.975±0.026 | 0.0351±0.003 | 28.4 | -4.847±0.0189 | 1.33 | -6.17 |

**Figure 2.29** Binding studies of Hakai (aa 106-206) with Cortactin using ITC. The tyrosine phosphorylated Cortactin was titrated against Hakai (aa 106–206) using ITC. The top panels show the heat release profiles after baseline correction and the lower panels indicate the binding isotherms for the interactions. The dissociation constant ($K_d$) and binding stoichiometry (N) are shown in the table.

As the ITC experiment confirmed the binding of the tyrosine phosphorylated Cortactin with Hakai (aa 106-206), we subsequently validated the idea that whether full length Hakai binds Src-phosphorylated cortactin, as well as the importance of Y482, Y485 and their flanking residues in this interaction. In addition to the WT proteins, phenylalanine substitution and alanine scan experiments were also performed on cortactin as described for E-cadherin. The results for cortactin mirror those obtained for E-cadherin. Like E-cadherin, cortactin interacts with Hakai only when phosphorylated by Src (Figure 2.30). Phenylalanine substitutions of the Src-phosphorylated tyrosine residues revealed that Y482 is the main tyrosine residue involved in Hakai binding (Figure 2.31). Furthermore, the acidic residues (E478, D480 and E486) surrounding Y482 contributed profoundly to the interaction between Hakai and E-cadherin, and significant contribution was also observed from S487 (Figure 2.32).

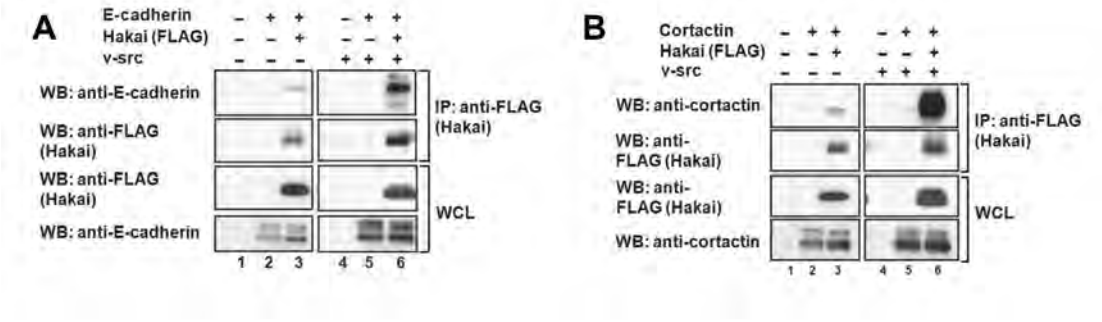



**Figure 2.30** In the similar manner as E-cadherin, Cortactin interacts with Hakai only when phosphorylated by Src. (A) E-cadherin and Hakai were co-transfected into HEK293 cells. Their interaction was analyzed through immunoprecipitation of FLAG-tagged Hakai. (B) Cortactin was co-transfected into HEK293 cells with Hakai. The interaction between cortactin and Hakai was compared with that in (A). The data clearly shows that cortactin binds Hakai in a pTyr-dependent manner in a way similar to that of E-cadherin.

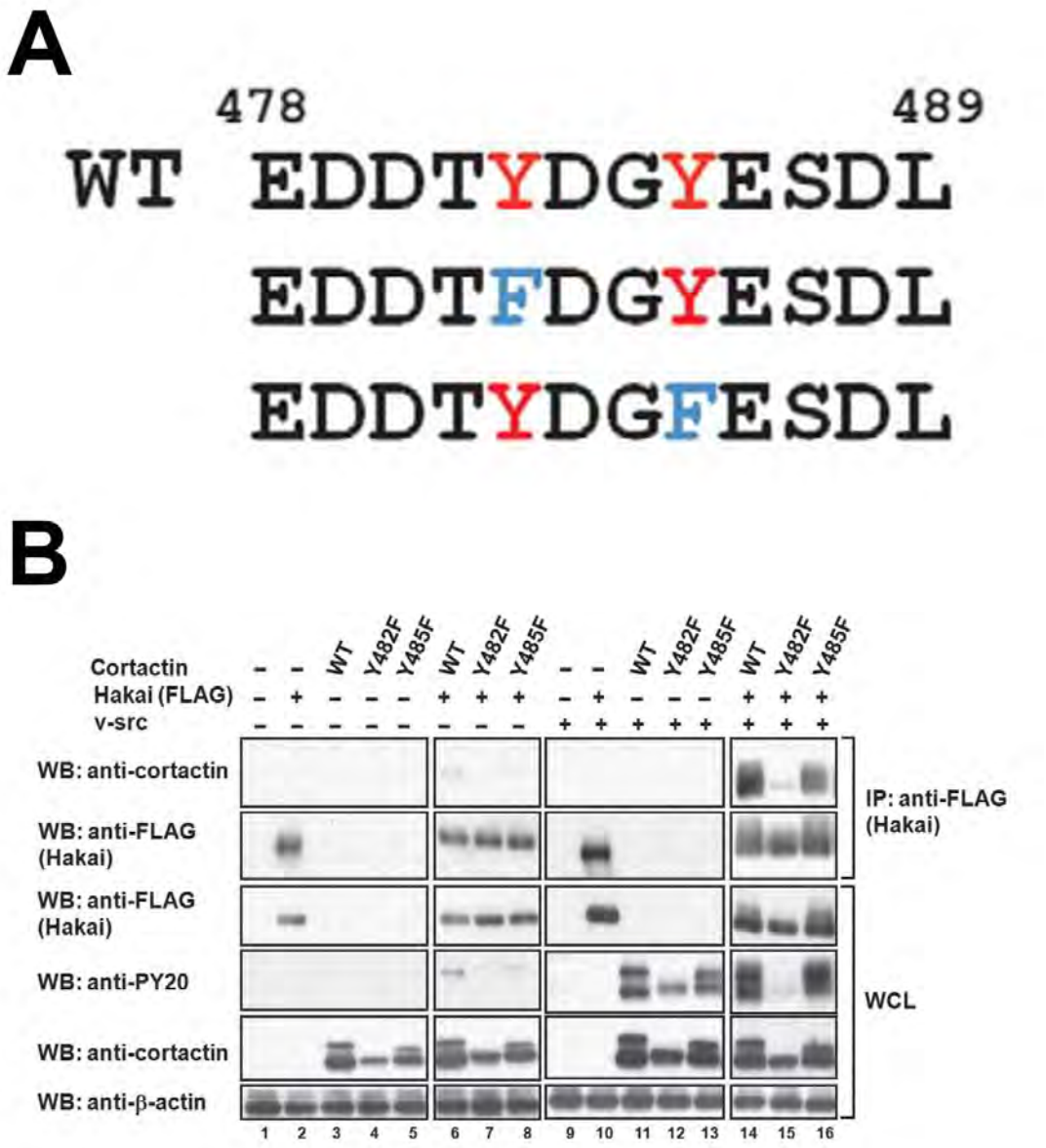



**Figure 2.31** Y482 is the main tyrosine residue of Cortactin that is involved in Hakai binding. (A) Y482 and Y485 (red) were separately substituted with phenylalanine (blue). (B) WT and mutated cortactin were co-transfected into HEK293 cells with Hakai, in the absence or presence of v-Src. The pTyr signal of Cortactin and its interaction with Hakai were analyzed.

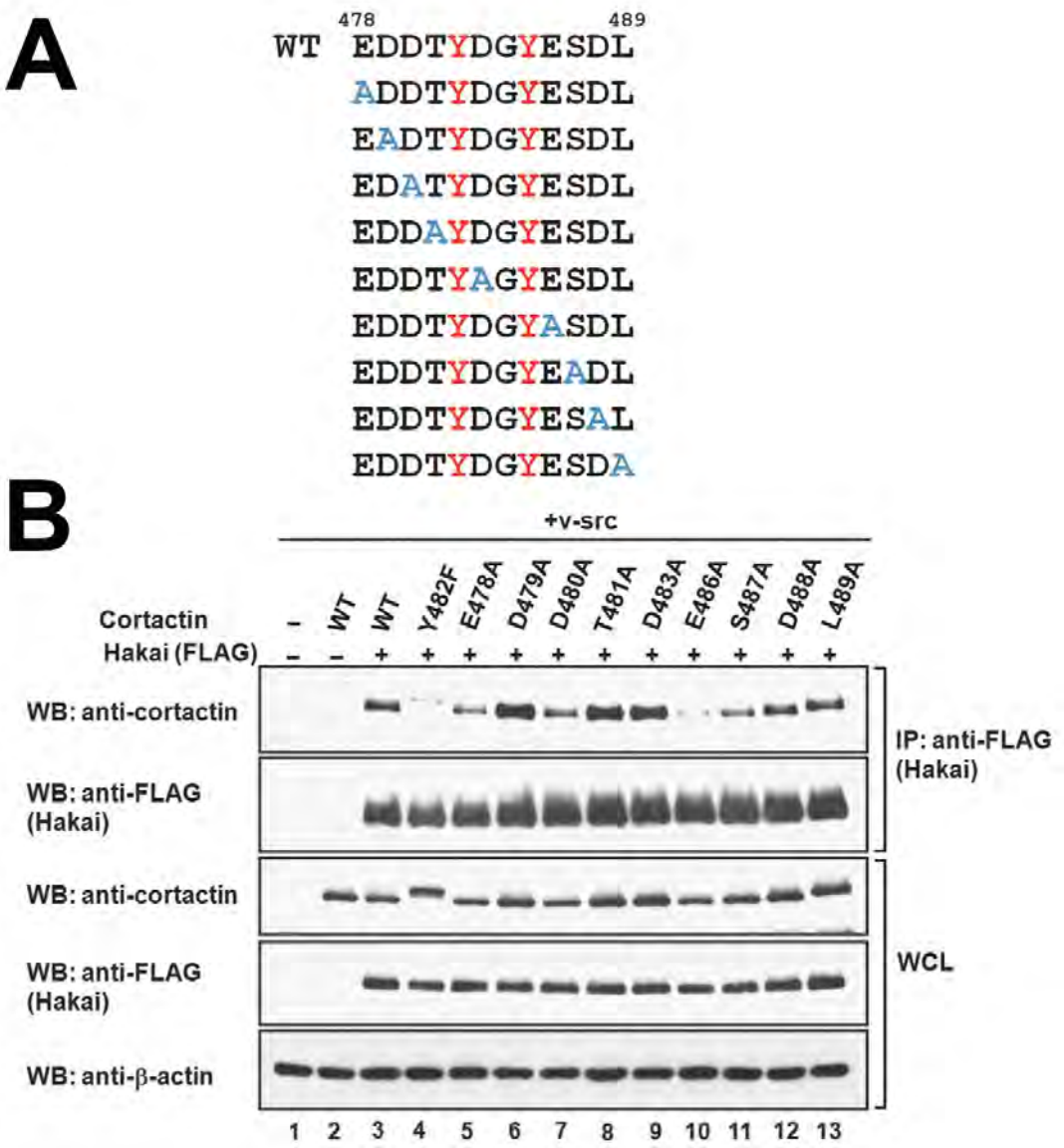



**Figure 2.32** Hakai recognizes acidic residues of Cortactin. (A) An alanine scan of cortactin aa 478–489. Each residue was substituted with alanine (blue). G484 was not mutated as glycine mutations affect the protein structure. (B) The cortactin mutants described in (A) were co-transfected into HEK293 cells with Hakai. The interaction between the cortactin mutants and Hakai was determined using immunoprecipitation.

The binding patterns of Hakai with E-cadherin and Cortactin suggested that Hakai recognizes Src phosphorylated tyrosine residue flanked by acidic amino acids. Thus, in order to further verify that Hakai targets pTyrs of Src substrates with surrounding acidic residues, another Src substrate was selected: DOK1, which also contains pTyrs with adjacent acidic groups (Luo et al., 2008). One such particular tyrosine residue

was found to be a primary phosphorylation site of Src (Luo et al., 2008) (Figure 2.33A). The results show that DOK1 interacts with Hakai (Figure 2.33B). Furthermore, DOK1 competed with endogenous Cortactin for Hakai, implying that DOK1 and Cortactin bind Hakai on the same site (Figure 2.33C). We also measured the binding affinity of the DOK1 pTyr motif with Hakai (aa 106-206) using ITC, which revealed that the binding is significantly strong with a $K_d$ value of 5.2μM (Figure 2.34).

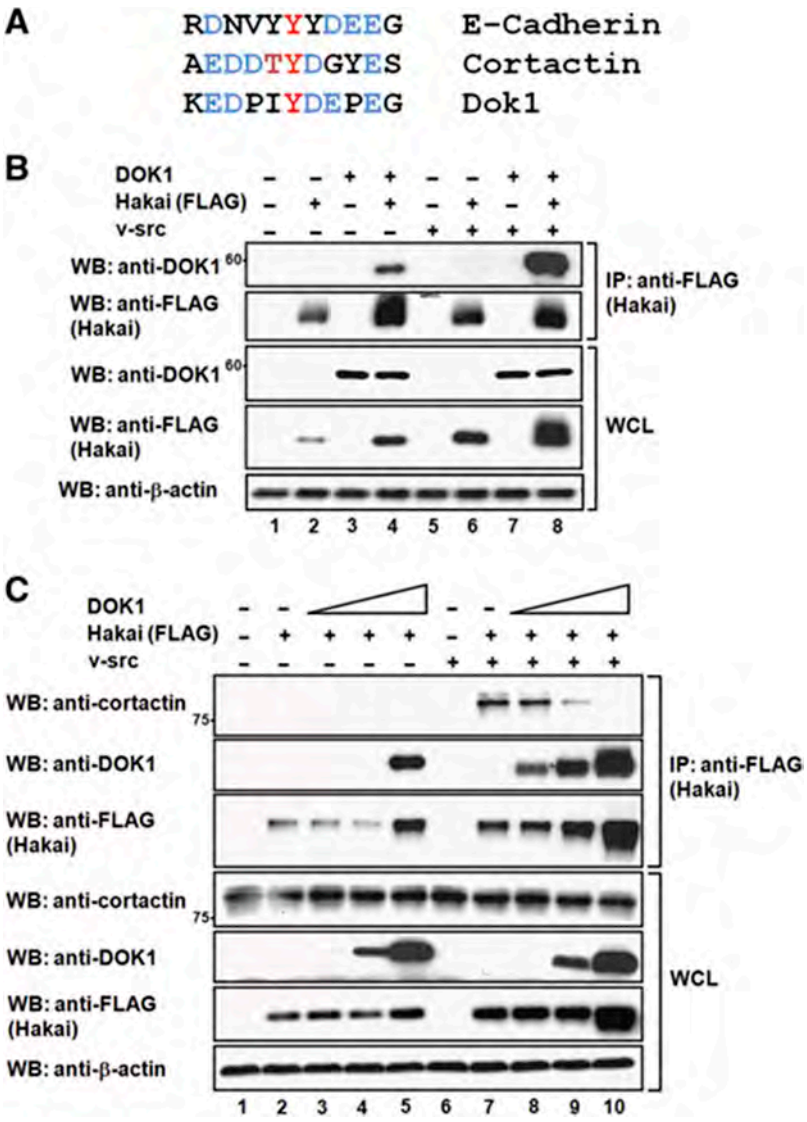
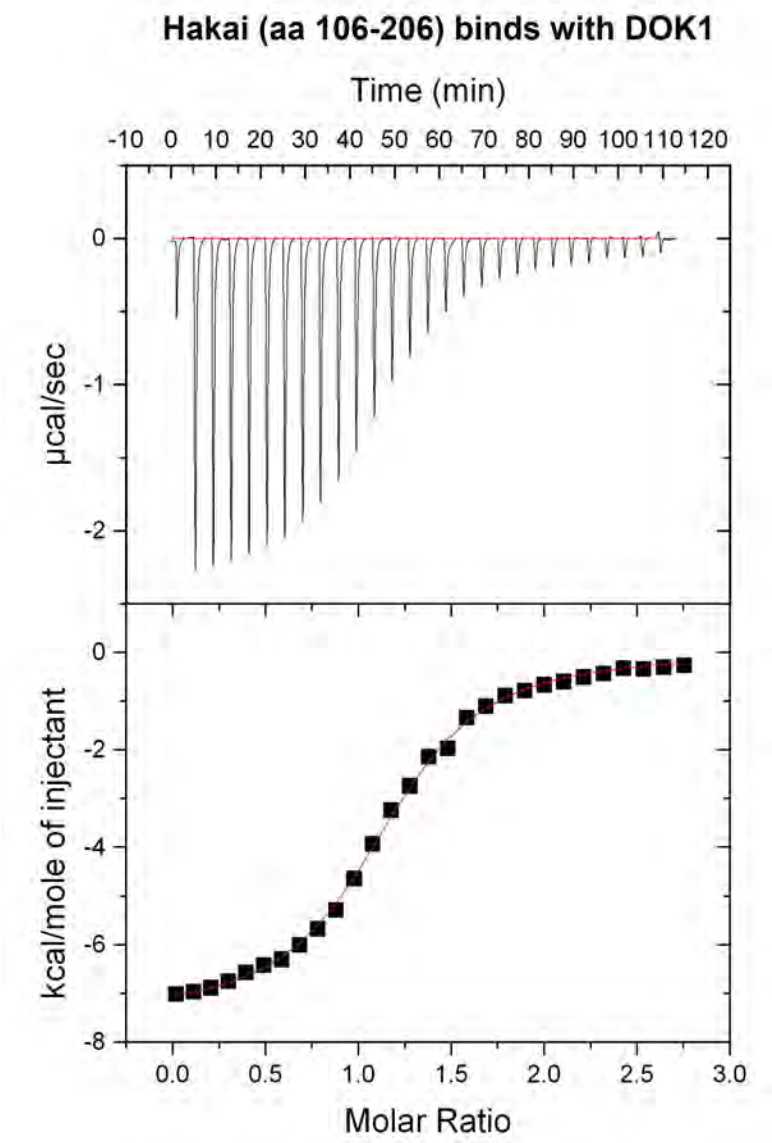



**Figure 2.33** DOK1 interacts with Hakai. (A) The sequence alignment of the different Src phosphorylation target sites in E-cadherin, cortactin and DOK1. The acidic

amino-acid residues flanking the phosphorylated tyrosine are shown in blue. (B) Hakai and DOK1 were overexpressed in HEK293 cells in the absence or presence of Src. FLAG immunoprecipitates were analysed for DOK1 interaction. (C) DOK1 was co-transfected into HEK293 cells with Hakai to study its competition with endogenous cortactin for binding to Hakai. FLAG immunoprecipitates were immunoblotted for cortactin.




| Peptide | Sequence | N | $K_a$ ($x10^6$ $M^{-1}$) | $K_d$ (µM) | ΔH (kcal/mol) | TΔS (kcal/mol) | ΔG (kcal/mol) |
|---|---|---|---|---|---|---|---|
| $DOK1^{356-366}$ | KEDPIpYDEPEG | 1.156±0.006 | 0.1923±0.008 | 5.2 | -7.445±0.048 | -0.237 | -7.208 |

**Figure 2.34** Binding studies of DOK1 with Hakai using ITC. The tyrosine phosphorylated DOK1 was titrated against Hakai (aa 106–206) using ITC. The top panels show the heat release profiles after baseline correction and the lower panels indicate the binding isotherms for the interactions. The dissociation constant ($K_d$) and binding stoichiometry (N) are shown in the table.

### 2.3.9 Determination of target-binding amino acids of Hakai

To identify the residues in Hakai within aa 106–206 necessary for its interaction with E-cadherin, 2D $^{1}H–^{15}N$-HSQC spectra of $^{15}N$-labelled Hakai (aa 106–206) were acquired in the absence and presence of a pTyr peptide derived from amino-acid residues 749–761 of E-cadherin. Based on the changes in the NMR spectrum of Hakai (aa 106–206), several residues underwent perturbations in chemical shift (Figure 2.35 and 2.36). A minimum criterion of a chemical shift difference $\Delta\delta_{p.p.m.} > 0.15$ p.p.m. was applied. Five residues were identified (Y176, H185, N187, H188 and R189), which resided in the Hakai minimum pTyr-binding domain, whereas a sixth residue (H127) was from the RING domain (Figure 2.37). However, the crystal structure shows that only four of these six residues reside in a close three-dimensional spatial proximity. While H127, Y176, H185 and R189 face the interior of the E-cadherin-binding site (Figure 2.37) and form a positively charged pocket, N187 and H188 face outward and are not part of the pocket (Figure 2.38).

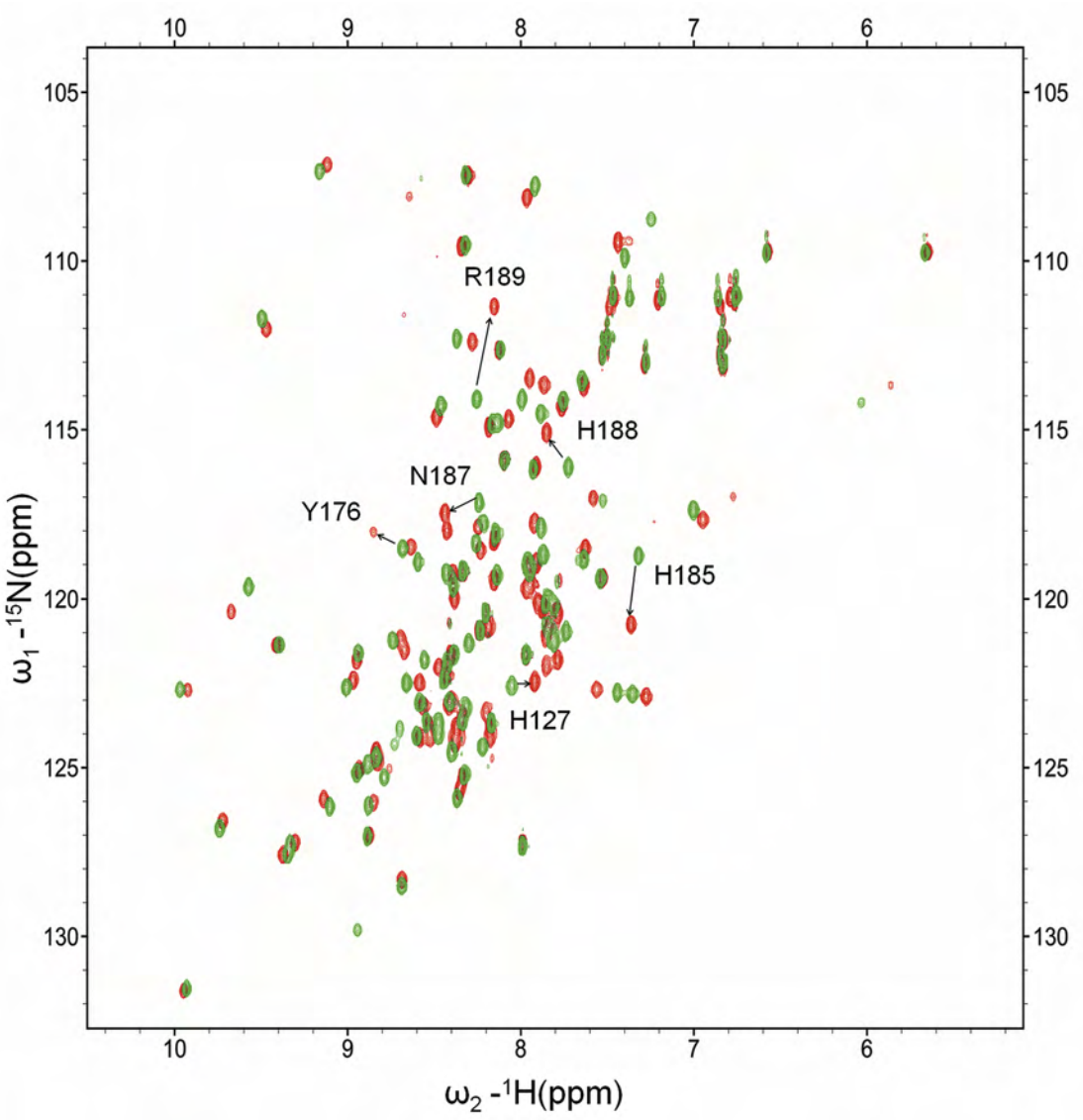



**Figure 2.35** NMR chemical shift perturbation experiment. An overlay of the $^{1}H$–$^{15}N$-HSQC spectra of Hakai (aa 106–206) in the absence (green) or the presence (red) of a tyrosine-phosphorylated E-cadherin peptide.

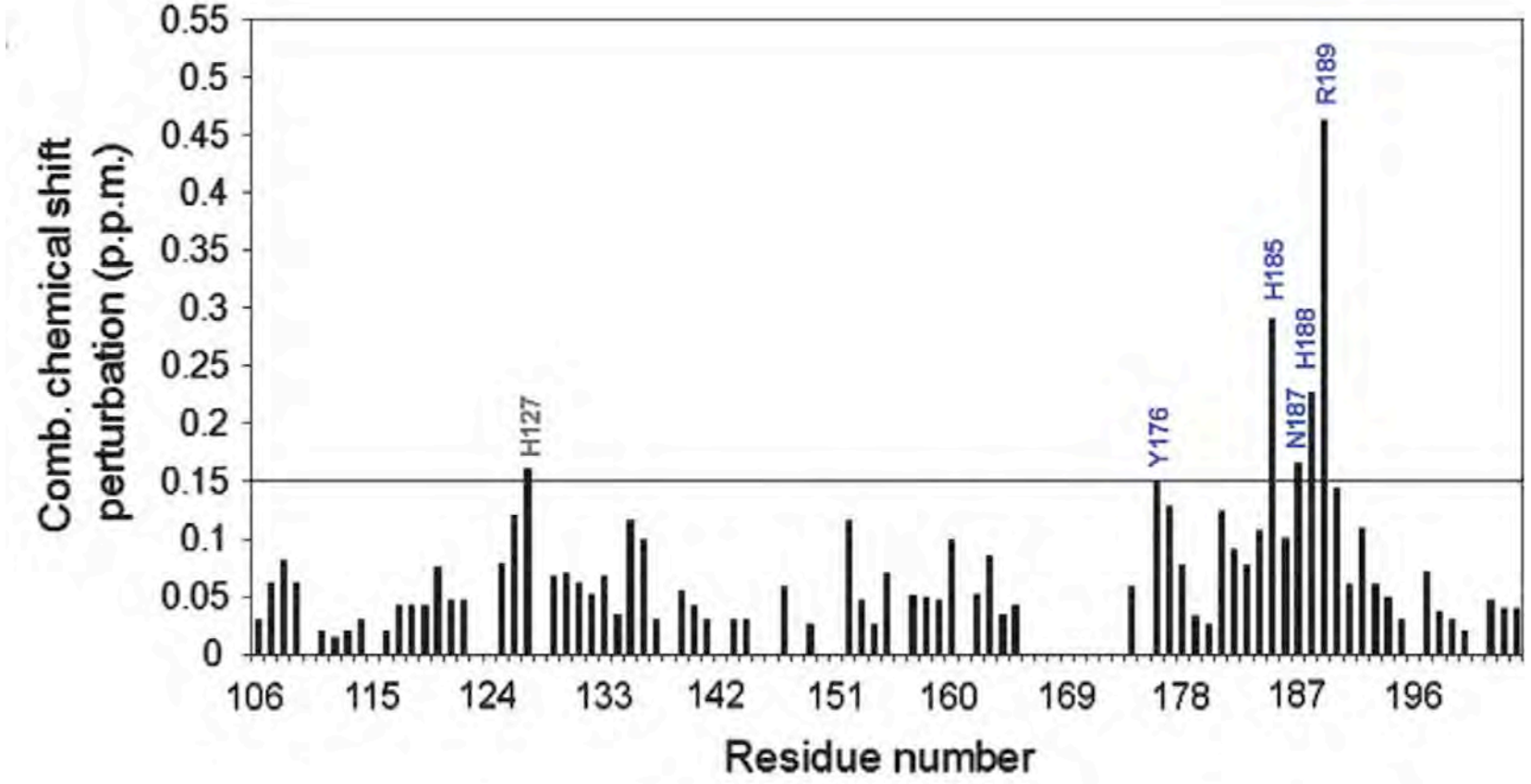



**Figure 2.36** Graphical representation of chemical shift perturbation. A graphical representation of the combined chemical shift perturbation (p.p.m.) plotted against all Hakai (aa 106–206) residues, with the cutoff at the combined chemical shift perturbation of 0.15 p.p.m.

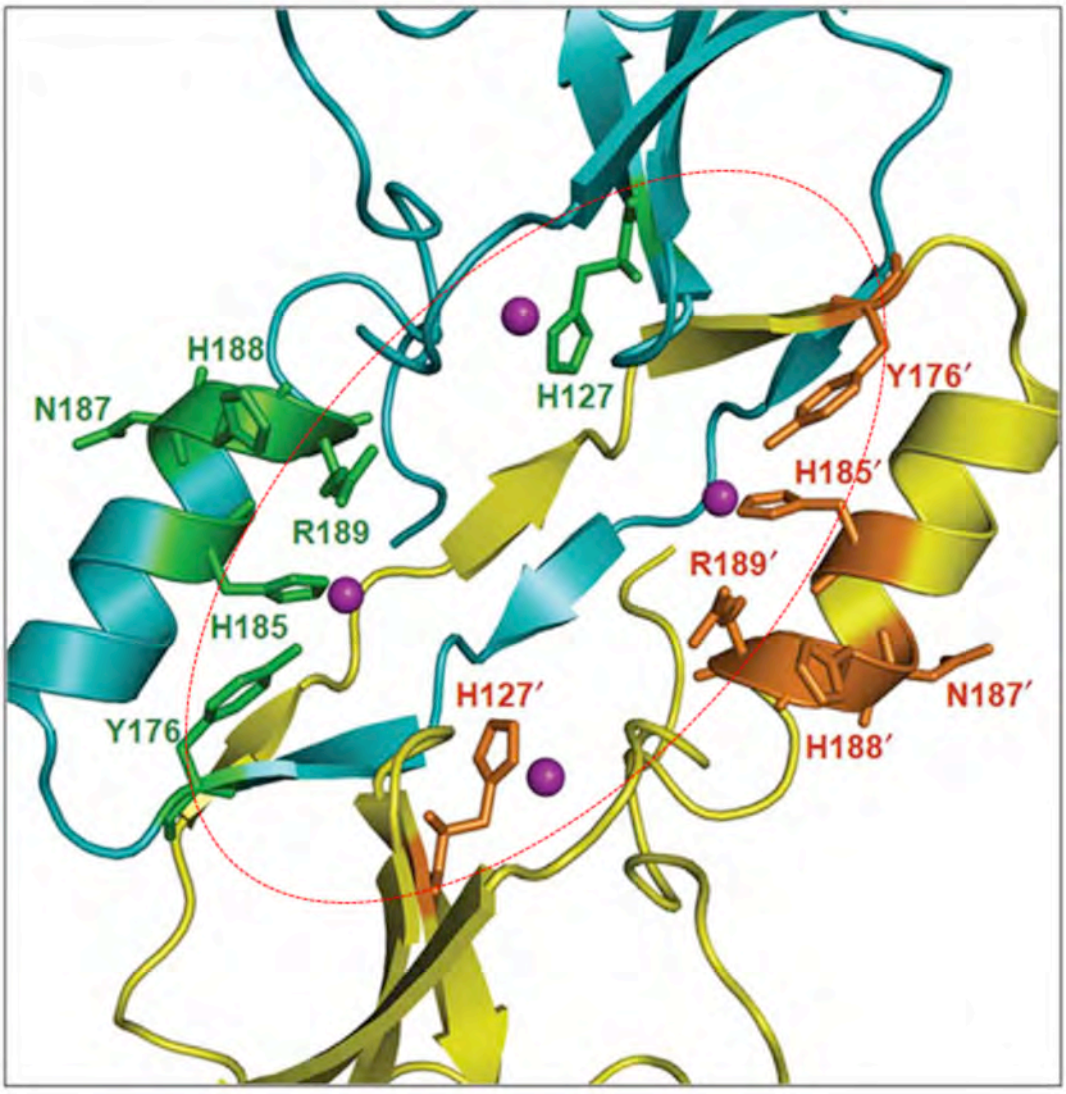



**Figure 2.37** Mapping of E-cadherin interacting residues on Hakai. The six potential E-cadherin-interacting residues in Hakai (aa 106–206) are highlighted as sticks in the ribbon representation of the crystal structure. The crystal structure shows that only four of these six residues reside in a close three-dimensional spatial proximity and face the interior of the E-cadherin-binding site, whereas, N187 and H188 face outward and are not part of the pocket.

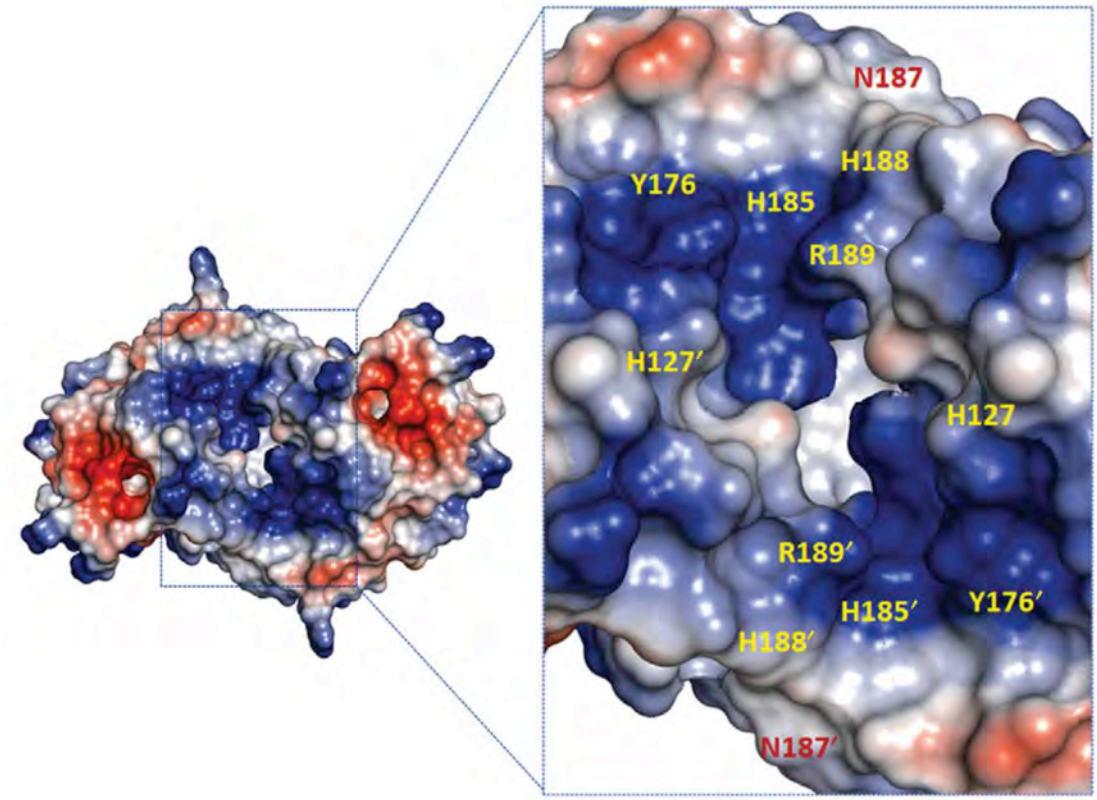



**Figure 2.38** A positively charged pTyr-binding pocket in Hakai. An electrostatic surface potential representation of Hakai (aa 106–206) shows that H127, Y176, H185 and R189 form part of the positively charged pocket.

These residues identified through *in vitro* peptide-domain binding assays were then further tested by expressing full-length point mutants in HEK293 cells. The immunoprecipitation results shown in Figure 2.39 indicate that the residues identified in the NMR analysis also abrogated binding when mutated, with the exception of residues N187 and H188. As expected, the required residues were situated on the interior of the target-binding domain, whereas the non-binding residues, N187 and H188, were on the exterior. Similar results were obtained with experiments using Cortactin (Figure 2.40A), demonstrating the importance of these Hakai residues. Furthermore, dimerization of Hakai is also required, as with E-cadherin, as cortactin was unable to bind to Hakai containing mutations to its zinc-coordinating residues (Figure 2.40B).

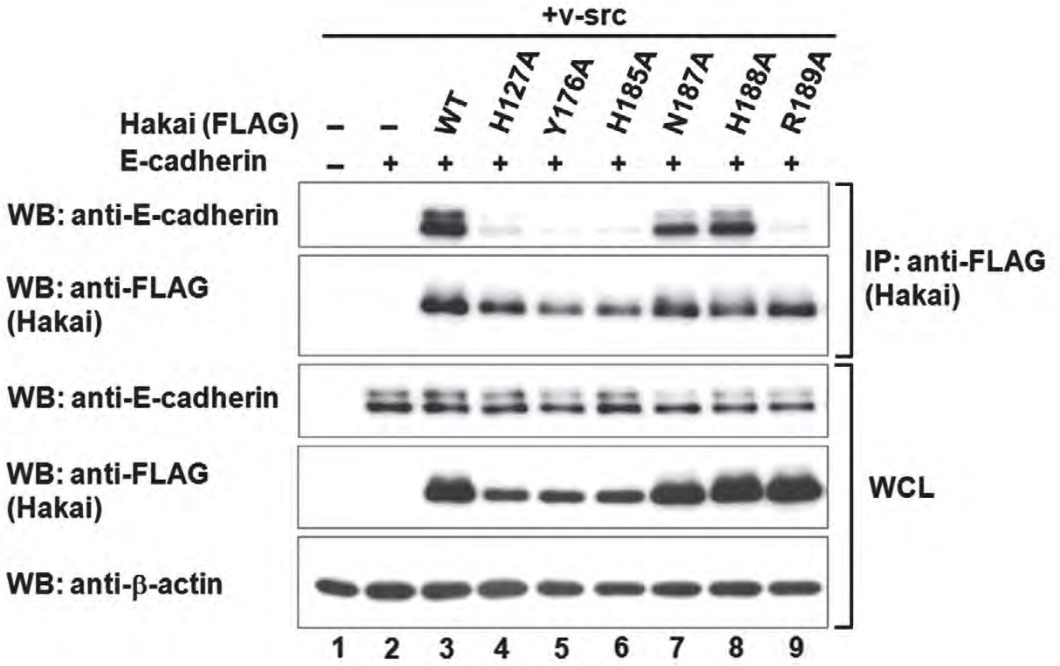



**Figure 2.39** Cell based validation using full length Hakai/E-cadherin binding. The interaction between E-cadherin and the Hakai mutants of the residues identified by NMR and shown in (Figure 2.38) was analyzed by immunoprecipitating FLAG-tagged Hakai.

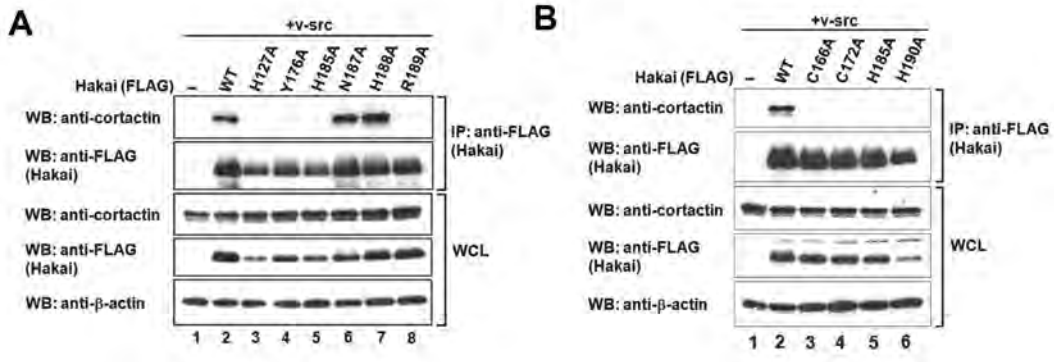



**Figure 2.40** Cell based interaction studies for validation of Cortactin binding residues of Hakai. As was the case with E-cadherin, the Hakai residues identified by NMR together with Hakai dimerization are important for binding with Cortactin. (A) HEK293 cells were transfected with the identified Hakai mutants, and their interaction with endogenous cortactin was studied. (B) Immunoprecipitates of either WT Hakai or the Hakai zinc-coordinating mutants were tested for interaction with endogenous cortactin.

Based on NMR data and mutational studies, a computational model was generated using Haddock (http://haddock.science.uu.nl/services/HADDOCK/haddock.php), which further confirms that the recognition sequence on E-cadherin would fit into the binding pocket as shown in Figure 2.41.

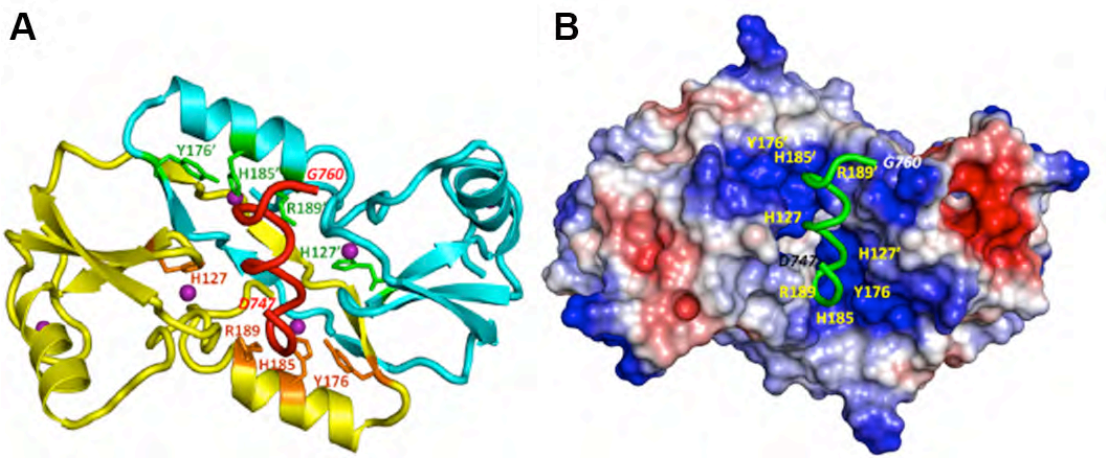



**Figure 2.41** A computational model based on the NMR data and mutational analyses shows that E-cadherin fits into the binding pocket of Hakai. (A) Ribbon representation showing the binding mechanism of Hakai (aa 106-206) with the E-Cadherin$^{747-760}$ peptide (B) Electrostatic surface potential representation of Hakai (aa 106-206) in the same orientation as shown in (A) indicating that the E-cadherin motif fits into the positively charged pocket of Hakai.

Based on the evidence obtained, it can be concluded that two Hakai monomers interact in an anti-parallel manner to form a dimer via the interlinked zinc-coordinating domain. This domain binds pTyrs flanked by acidic amino acids in Src substrates. The target-binding domain resulting from this dimerization process represents the functional Hakai phosphotyrosine-binding domain, henceforth referred to as the HYB (**H**akai p**Y**-**b**inding) domain (Figure 2.42).

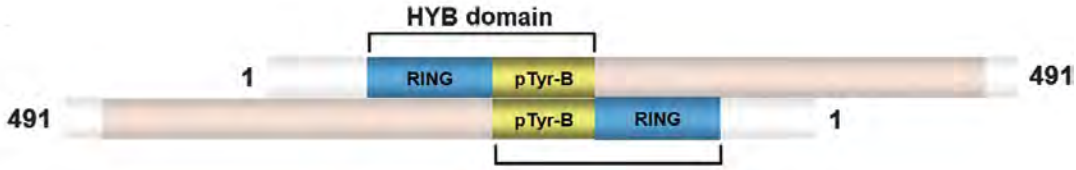



**Figure 2.42** A schematic representation of the Hakai dimer and the HYB domain.

### 2.3.10 The HYB domain in other proteins

We next investigated whether the HYB domain is found in other proteins. Literature and database searches revealed that the testis-specific ubiquitin E3 ligase ZNF645 exhibited high sequence homology with Hakai (Liu et al., 2010), as shown in Figure 2.43A. We therefore questioned whether ZNF645 could also interact with E-cadherin and cortactin. The results in Figure 2.43B show that ZNF645 bound to v-Src-phosphorylated E-cadherin but not to cortactin (Figure 2.43C). This result implies that although there is significant homology between Hakai and ZNF645, they are likely to have their own sets of targets due to the differences in their sequences between the key zinc-coordinating residues.

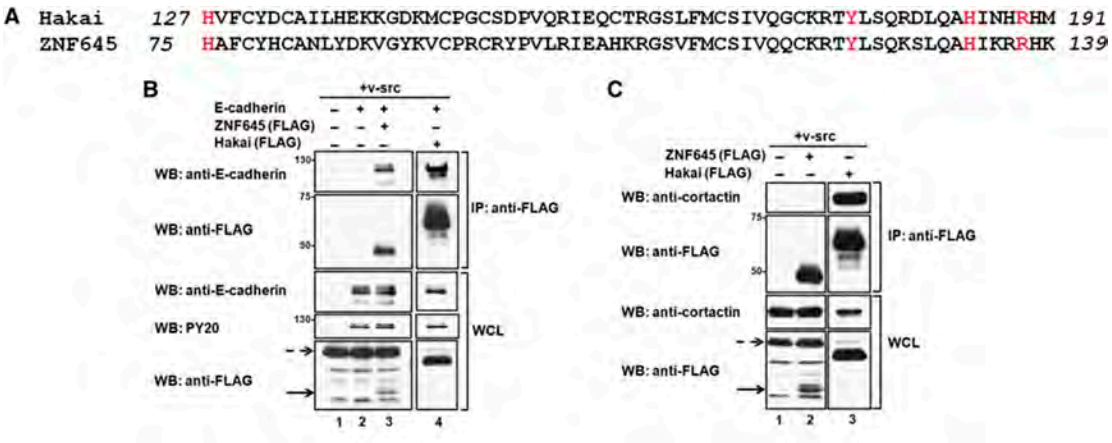



**Figure 2.43** The HYB domain in other proteins. (A) A comparison of the Hakai protein from amino-acid residues 127–191 and the equivalent sequence in ZNF645. (B) E-cadherin and ZNF645 were analysed for their interaction using

immunoprecipitation. Hakai was used as a positive control. The dotted arrow indicates a non-specific band; the solid arrow indicates the ZNF645 band. (C) ZNF645 was overexpressed in HEK293 cells and its interaction with endogenous cortactin was analyzed using immunoprecipitation.

Based on the key amino-acid residues involved in zinc coordination and binding in HYB, we searched the NCBI database to analyze gene origins and protein homologies (Table 2.6). Two interesting results emerged. First, a comparison of the species distribution of the Hakai and ZNF645 gene products indicated that the latter, found only in primates, is most likely a recent copy of the former. Second, ZNF645 is an intronless gene, implying that it is a retrotransposed copy of Hakai.

| | Species | HVFCYDCAILHEKKGDKMCPGCSDPVQRIEQCTRGSLFMCSIVQGCKRTYLSQRDLQAHINHRH |
|---|---|---|
| HAKAI | Homo sapiens (Human) | HVFCYDCAILHEKKGDKMCPGCSDPVQRIEQCTRGSLFMCSIVQGCKRTYLSQRDLQAHINHRH |
| | Bos taurus (Cattle) | HVFCYDCAILHEKKGDKMCPGCSDPVQRIEQCTRGSLFMCSIVQGCKRTYLSQRDLQAHINHRH |
| | Macaca mulatta (Rhesus monkey) | HVFCYDCAILHEKKGDKMCPGCSDPVQRIEQCTRGSLFMCSIVQGCKRTYLSQRDLQAHINHRH |
| | Rattus norvegicus (Norway rat) | HVFCYDCAILHEKKGDKMCPGCSDPVQRIEQCTRGSLFMCSIVQGCKRTYLSQRDLQAHINHRH |
| | Ailuropoda melanoleuca (Giant panda) | HVFCYDCAILHEKKGDKMCPGCSDPVQRIEQCTRGSLFMCSIVQGCKRTYLSQRDLQAHINHRH |
| | Mus musculus (House mouse) | HVFCYDCAILHEKKGDKMCPGCSDPVQRIEQCTRGSLFMCSIVQGCKRTYLSQRDLQAHINHRH |
| | Equus caballus (Horse) | HVFCYDCAILHEKKGDKMCPGCSDPVQRIEQCTRGSLFMCSIVQGCKRTYLSQRDLQAHINHRH |
| | Canis lupus familiaris (Dog) | HVFCYDCAILHEKKGDKMCPGCSDPVQRIEQCTRGSLFMCSIVQGCKRTYLSQRDLQAHINHRH |
| | Pan troglodytes (Chimpanzee) | HVFCYDCAILHEKKGDKMCPGCSDPVQRIEQCTRGSLFMCSIVQGCKRTYLSQRDLQAHINHRH |
| | Gallus gallus (Chicken) | HVFCYDCAILHEKKGDKMCPGCNEPVQRIEQCVRGSLFMCSIVQGCKRTYLSQRDLQAHINHRH |
| | Xenopus laevis (African clawed frog) | HVFCYDCALMHEKKADKLCPGTLVEESTDTFKRMSCNDPVQRIEQCARGSLFMCSIVQGCKRTYLSQRDLQAHINHRH |
| | Danio rerio (Zebrafish) | HVFCYDCAVYYEKKCDKMCPGLSLYSCTDPVQRIEQCQRGSLFMCNIVQGCKRTYLSQRDLQAHINHRH |
| | Salmo salar (Atlantic salmon) | HVFCYDCALLHEKKMEKMCPGLTLYSCTDPVQRIEQCLRGLLYMCSIVPGCKRTYLSQRDLQAHVNHRH |
| | Harpegnathos saltator (Jerdon's jumping ant) | HVFCLSCA----KREDKVCPRCMEKVSRVEQTGLGTVFMCTHGGTRYGNTGCRRTYLSQRDLQAHINHRH |
| | Camponotus floridanus (Carpenter ant) | HVFCLSCA----KREDKVCPRCMEKVSRVEQTGLGTVFMCTHGGTRYGNAGCRRTYLSQRDLQAHINHRH |
| | Tetraodon nigroviridis (Green pufferfish) | HVFCYDCALLHEKKGEKMCPGLTLYNCTDPVQRIEQCQRGSLYMCSVVPGCKRTYLSQRDLQAHVNHRH |
| | Anopheles darlingi (Mosquito) | HVFCLRCA---RSETLKMCPRCKEKVVRVEQTALGTVFMCTHGGTRYGNTGCRRTYLSQRDLQAHINHRH |
| | Drosophila grimshawi (Fruit fly) | HVFCLKCA---RAEPIKCCPRCNDKVLRVEQSGLGTVFMCTHGGSRYGSTGCRRTYLSQRDLQAHINHRH |
| | Tribolium castaneum (Red flour beetle) | HVFCLSCG----KQEQKQCPRCLEKVSRVEQTGLGTVFMCTHGGTRYGSSGCRRTYLSHRDLQAHINHRH |
| | Papilio xuthus (Asian swallowtail butterfly) | HVFCLSCA----RSDHTHCPRCREKVLRVEQTGLGTVFMCTHSGTRYGNTGCRRTYLSQRDLQAHINHRH |
| | Hydra magnipapillata (Freshwater hydrozoan) | HVFCLSCA----ENSNGECVRCEERIDRIEPATIGQIFVCSFGGNRNITSGCRRSYLSQRDLIAHIRHRH |

| | Species | HAFCYHCANLYDKVGYKVCPRCRYPVLRIEAHKRGSVFMCSIVQQCKRTYLSQKSLQAHIKRRH |
|---|---|---|
| ZNF645 | Homo sapiens (Human) | HAFCYHCANLYDKVGYKVCPRCRYPVLRIEAHKRGSVFMCSIVQQCKRTYLSQKSLQAHIKRRH |
| | Macaca fascicularis (Cynomolgus Monkey) | HAFCYNCANLYDKIGYKICPRCSYPVLRIEEHKRGSVFMCSVVQGCKRTYLSQKSLQAHIKRRH |
| | Macaca mulatta (Rhesus monkey) | HAFCYNCANLYDKIGYKICPRCSYPVLRIEEHKRGSVFMCSVVQGCKRTYLSQKSLQAHIKRRH |
| | Pongo abelii (Sumatran orangutan) | HAFCYDCANLDDKIGYKICPRCRYPVLRIEEHKRGSVFMCSVVPQCKRTYLSQKSLQAHIKRRH |

**Table 2.6** List of proteins with sequences that potentially match the region of Hakai from amino acid residues 127 to 189.

Further database searches based on the conserved zinc-coordinating cysteine and histidine residues within the HYB domain showed that a similar series of residues is present in LNX1 and LNX2 (Figure 2.44). This implies that the HYB domain may be

distributed in other proteins, although the latter observation requires experimental confirmation.

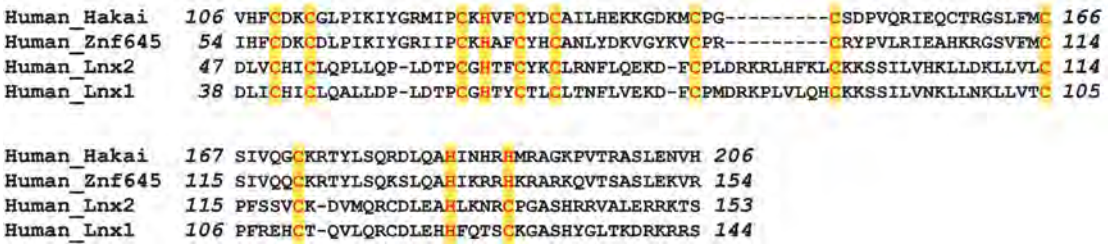



**Figure 2.44** Sequence alignment of LNX1 and LNX2 with Hakai and ZNF645 based on the conserved zinc-coordinating residues from Hakai aa 106–206.

### 2.3.11 Novel structure of the HYB domain

The structures of all the five pTyr-binding domains that have been discovered to date, which also includes the two idiosyncratic pTyr-binding domains (C2 domain of PKCδ and PKM2), are illustrated in Figure 2.45A and B. All of the domains, except for the HYB domain, are contained within one monomer. The HYB domain consists of a pair of monomers arranged in an anti-parallel configuration and is composed of two RING and two atypical zinc-coordinating domains. From this comparison, it is apparent that all five of these pTyr domains have completely different structures, with different strategies to recognize tyrosine phosphorylation.

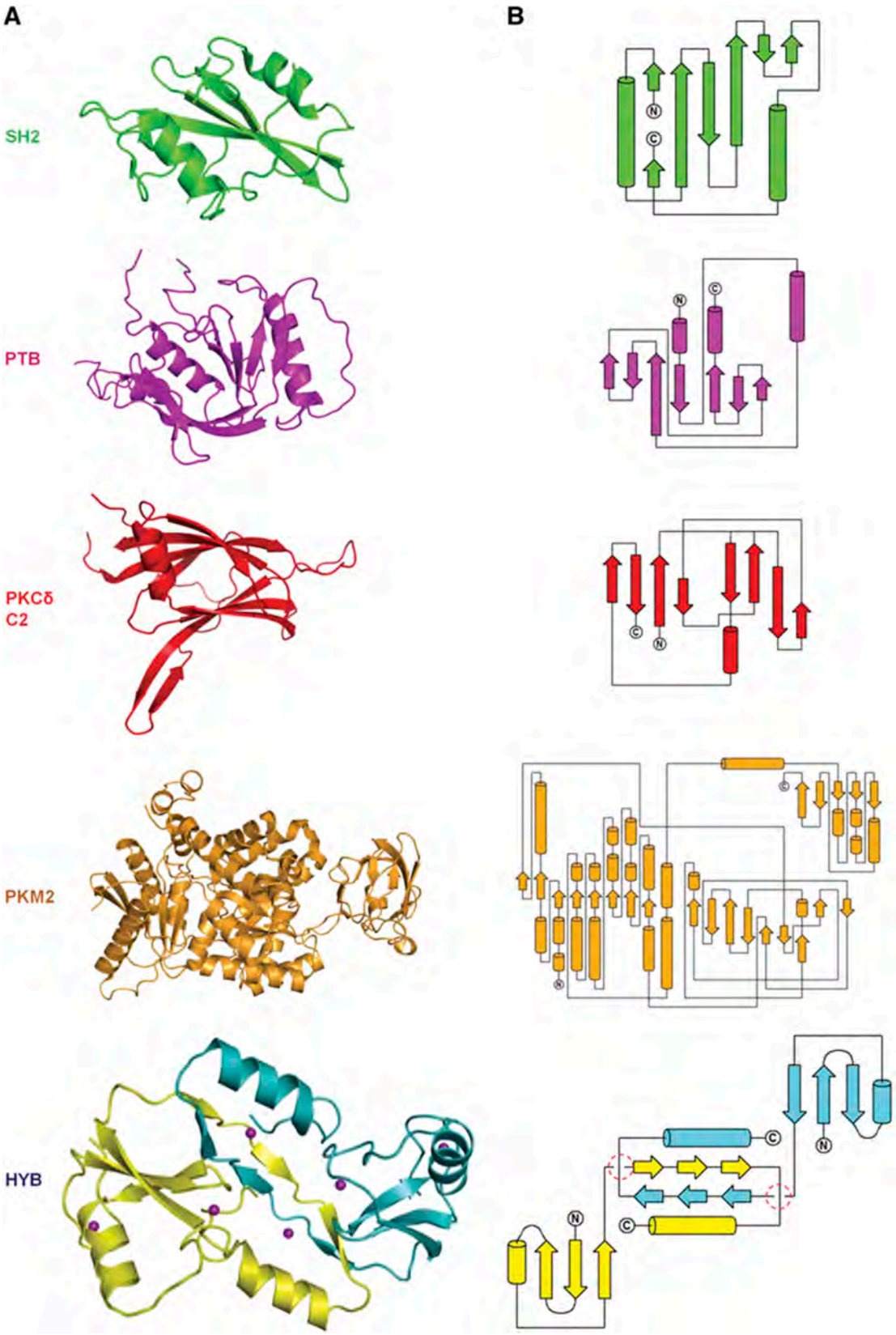



**Figure 2.45** Novel structure of the HYB domain. (A) Representative structures of SH2 (PDB code 1SHB), PTB (PDB code 1SHC), PKCd C2 (PDB code 1YRK) and PKM2 (PDB code 3BJF) in ligand-free forms are compared with the HYB domain. (B) The corresponding topologies of the domains in (A).

## 2.4 Discussion

Zinc fingers (ZnFs) consist of a diverse group of protein domains that are able to bind to zinc ions. Initially thought to bind only nucleic acids, ZnFs were later shown to mediate protein– protein interactions (Gamsjaeger et al., 2007; Klug, 2010). The common structural feature in the different ZnF classes, where determined, is that the zinc-coordinating cysteine and histidine residues are all situated on the same monomer. The evidence in this study has shown that Hakai, in addition to having a typical RING domain, possesses an atypical ZnF with a novel structural fold. In contrast to previously reported structures, the formation of this atypical ZnF requires amino acid residues from two intertwining Hakai monomers. Three of the four residues are contributed by one monomer, whereas one cysteine residue comes from another monomer, implying that the formation of this ZnF is accompanied by dimerization. Although this particular ZnF is novel by itself, it is made more intriguing by the fact that this domain lies within a pTyr-binding region. The results from our study demonstrate that in conjunction with the RING domain of Hakai, dimerization of Hakai through the formation of the atypical ZnF creates a novel pTyr-binding domain, called the HYB domain. Originally postulated to recognize multiple consecutive phosphotyrosines, it was instead found that the HYB domain recognizes a single phosphotyrosine residue flanked by acidic amino acids. Interestingly, this may indicate a bias toward Src substrates because Src kinase has a preferred, but not strict, recognition sequence of Y(p)-E-E-I/V (Yeatman, 2004).

The zinc-coordinating residues of each Hakai monomer consist of a conserved series of nine cysteine and three histidine residues that facilitate Hakai dimerization. Based

on sequence analysis, LNX1 and LNX2 also harbor such a sequence motif, although the last conserved histidine is replaced by cysteine. While the full structures of the LNX proteins are currently unknown, it is possible that, based on sequence homology, they also contain the HYB domain. As the HYB domain contains a RING domain, its occurrence would be predominantly in E3 ubiquitin ligases, which has implications in the way these proteins recognize their targets and effect downstream signaling.

Although the zinc-coordinating residues within the ZnFs of the HYB domain are critical, the protein folding of the primary sequences surrounding the cysteine and histidine residues are also important in defining its characteristics. This chelation-derived folding provides the necessary three-dimensional structure while specific amino acids within the sequences between the zinc-coordinating residues are instrumental in recognizing and binding phosphorylated tyrosines of target proteins. In the case of Hakai, four residues were identified through NMR studies to bind to E-cadherin, H127, Y176, H185 and R189. It is noteworthy that there is strict conservation of these residues between Hakai and ZNF645. On the other hand, the differences in their binding targets are most likely dictated by amino acids that are not common to both proteins, although this aspect requires further study. It would also be interesting to observe if LNX1 or LNX2 binds phosphorylated tyrosine residues. Based on sequence homology, both LNX1 and LNX2 harbor most of the zinc-coordinating residues of the HYB domain. However, only LNX2 has the equivalent of R189 of Hakai. Neither of the LNX proteins has the equivalent of Y176. It is possible that these two proteins bind a different set of tyrosine-phosphorylated proteins by utilizing different residues within their sequences. Thus while the basic sequence required to provide a unique pTyr-binding structure is apparent within the four

proteins, a number of questions regarding the exact pTyr-binding requirements of these proteins remain unanswered.

In conclusion, we have identified a novel HYB domain, made up of two structurally atypical ZnF domains, which recognize a previously unidentified phosphotyrosine motif. Our structural data indicate that the HYB domain forms a positively charged pocket, implying that it represents a highly suitable drug target. As dysfunctional E3 ubiquitin ligases are often associated with diseases (Burger et al., 2006; Lohr et al., 2010), the presence of the HYB domain in these proteins offers a specific target for directed therapies.

To summarize, this chapter described the discovery of a novel pTyr-binding domain, the HYB domain, which bears a dimeric fold. This is in contrast to the existing pTyr-binding domains, typified by SH2 and PTB domains, which generally function as monomer. In the next chapter, we have designed a C-terminal deletion mutant of the HYB domain, $HYB^{\Delta C}$, which exists as a monomer in solution. Further, we have studied the pTyr-binding property of $HYB^{\Delta C}$.

# CHAPTER III

## Dimeric switch of Hakai-truncated monomers during substrate recognition: insights from solution studies and NMR Structure

## 3.1 Introduction

Phosphotyrosine (pTyr) binding domains are key determinants of specificity and selectivity in many signal transduction pathways and act by integrating pTyr signals from upstream kinases to downstream effectors that regulate the complex physiology of eukaryotic cells (Deribe et al., 2010; Hunter, 2009; Yaffe, 2002). The SH2 domain was the first pTyr-binding domain to be discovered (DeClue et al., 1987; Sadowski et al., 1986), and has since been extensively characterized (Filippakopoulos et al., 2009; Forman-Kay and Pawson, 1999; Liu et al., 2006; Pawson and Nash, 2003; Songyang et al., 1993; Yaffe, 2002). The second domain to be identified with a capacity to bind tyrosine residues was the Phosphotyrosine Binding (PTB) domain (Kavanaugh and Williams, 1994). Subsequently, idiosyncratic pTyr-binding domains have been observed in the C2 domain of Protein Kinase C δ (Benes et al., 2005) and Pyruvate Kinase M2 (Christofk et al., 2008), and we recently reported the existence of a third pTyr-binding domain in Hakai and coined it the HYB domain for Hakai phosphotyrosine binding (Mukherjee et al., 2012).

Hakai is an E3 ubiquitin ligase first noted for its role in regulating E-cadherin expression and disrupting cell-cell contacts in epithelial tissues (Fujita et al., 2002; Pece and Gutkind, 2002). Subsequent work has identified an upregulation of Hakai in human colon and gastric adenocarcinomas (Abella et al., 2012; Figueroa et al., 2009). The HYB domain interacts with other Src substrates, such as Cortactin, a structural protein involved in coordinating actin rearrangement during cell movement (Wu et al., 1991), and DOK1, a scaffolding protein that assists in the assembly of signalling complexes (Mashima et al., 2009). Both of these proteins have important functional

contributions in the progression of cancer (Noh et al., 2013; Siouda et al., 2012), and both are regulated by the activity of Hakai (Mukherjee et al., 2012). The novel features of the HYB domain and its interaction with various key molecules may indicate a physiologically important role for Hakai in cancer.

The HYB domain comprises amino acids 106-206, and forms an atypical, zinc-coordinated homodimer in an antiparallel, intertwined configuration, utilizing residues from the phosphotyrosine-binding region of two Hakai monomers. The dimeric nature of the domain configures the formation of a phosphotyrosine-binding pocket that recognizes specific phosphorylated tyrosine residues and flanking acidic amino acids of its substrates (Mukherjee et al., 2012). The C-terminal region, which harbors the atypical zinc-coordination motif and key residues involved in the pTyr interaction, plays a key role in the dimerization observed in the HYB domain. In the present study, we investigated the consequences of deleting residues from the C-terminus of Hakai whilst maintaining the key residues involved in the pTyr interaction and the atypical zinc-coordination of the E3 ligase. Through this structural truncation, we identified a C-terminal deletion mutant, Hakai (aa 106-194), herein referred to as $HYB^{\Delta C}$, which exists as a monomer in solution and flips to a dimeric arrangement in the presence of a pTyr substrate, with the incoming substrate inducing the conformational changes required to initiate dimerization of $HYB^{\Delta C}$ monomers. This monomeric to dimeric switch of $HYB^{\Delta C}$ in the presence of the phosphorylated substrate was further validated by biophysical and solution studies. Taken together, our results suggest that the dimerization of Hakai pTyr-binding domain is a unique prerequisite for substrate binding.

## 3.2 Materials and Methods

### 3.2.1 Protein expression and purification

Inserts corresponding to Hakai (amino acids 106-194) have cloned into pGEX-6P-1 vector (GE Healthcare) using *BamHI* and *SalI* as the restriction sites and expressed as a GST-fusion protein in *E. coli* BL21 (DE3) cells. The cells were cultured in Luria Broth medium at 37°C until the $A_{600}$ nm reached 0.6–0.7, IPTG and zinc sulphate were then added at concentrations of 0.15mM and 50μM, respectively, and growth continued for around 20 hours at 15°C. Cells were harvested by centrifugation (8000 × g, 10 min, 4°C), and the pellet was resuspended in lysis buffer (50 mM Bis-Tris pH 6.5, 300mM NaCl, 10μM zinc chloride, 5% Glycerol, 0.5% Triton-X 100, 2mM DTT, 1mM phenylmethylsulfonylfluoride (PMSF) and homogenized using a French Press Cell Disrupter (Thermo Electron Corporation). Subsequently, the cell lysate was centrifuged at 18,000 rpm for 30 min at 4°C (JA-25.50 fixed angle rotor, Beckman Coulter centrifuge) and the supernatant was allowed to bind to glutathione-sepharose resin (GE Healthcare) for 2-4 hours at 4°C and was subsequently washed with a buffer containing 50 mM Bis-Tris pH 6.5, 300mM NaCl, 5% glycerol, 2mM DTT. After the washing step, an overnight on-column cleavage at 4°C with GST-PreScission Protease (GE Healthcare) was performed to remove the GST tag. A major portion of GST and GST-PreScission Protease remained bound to the glutathione-sepharose resin and the flow-through containing the partially purified, untagged proteins were further purified using a Superdex 75 size exclusion column (GE Healthcare) equilibrated with a buffer containing 50mM sodium phosphate buffer pH 6.5 and 5mM DTT.

### 3.2.2 NMR spectroscopy

All NMR experiments were carried out at 25 °C on a Bruker Avance 800 MHz spectrometer equipped with a TXI cryogenic probe. $^{1}H$, $^{13}C$, and $^{15}N$ resonance assignments were performed by measuring the three-dimensional HNCACB, three-dimensional CBCA(CO)NH (Bax and Grzesiek, 1993), and three-dimensional CCH-TOCSY (Fesik et al., 1990) spectra. Inter-proton distance restraints for structural calculation were obtained from three-dimensional $^{13}C$-edited NOESY-HSQC, three-dimensional $^{15}N$-edited NOESY-HSQC, and two-dimensional NOESY spectra using a 100-ms mixing time. $^{1}D_{NH}$ residue dipolar couplings were measured using the IPAP method (Ottiger et al., 1998). The residue dipolar coupling values were obtained by subtracting the reference value in isotropic solution. Two- and three-dimensional NMR spectra were processed using the NMRPipe program (Delaglio et al., 1995), and data analysis was performed with the help of the Sparky program (Goddard and Kneller, 2004).

The structure was calculated using the Xplor-NIH 2.24 software package (Schwieters et al., 2003). A total of 1014 NOE distance restraints, 21 hydrogen bonds, and 110 dihedral angle restraints were predicted by the TALOS program (Cornilescu et al., 1999). Residue dipolar couplings were used in the final cooling stage. 100 structures were calculated using 30,000 steps of simulated annealing, and a final ensemble of 20 lowest energy structures was selected for figure preparation.

### 3.2.3 Size-exclusion chromatography

$HYB^{\Delta C}$ protein samples, with concentrations at $A_{280\ nm}$ of 1.0, in the absence or presence of phosphorylated E-cadherin$^{749–761}$ peptide, was applied to a Superdex 75 size-exclusion column (GE Healthcare) equilibrated using in buffer containing 50mM sodium phosphate pH 6.5 and 5mM DTT and pre-calibrated using a gel-filtration standard (Bio-Rad).

### 3.2.4 Circular Dichroism spectrometry

Far UV spectra (260–190 nm) of $HYB^{\Delta C}$ and HYB were measured using a Jasco J-810 spectropolarimeter in phosphate buffer (pH 6.5) at room temperature using a 0.1-cm path length and stoppered cuvettes. Six scans were recorded, averaged and the baseline subtracted.

### 3.2.5 Dynamic light scattering

Dynamic light scattering studies were carried out on a DynaPro Light Scattering instrument (Protein Solutions, USA) with protein concentrations at $A_{280\ nm}$ of 1.0, in buffer containing 50mM sodium phosphate pH 6.5 and 5mM DTT.

### 3.2.6 Isothermal Titration Calorimetry

Phosphorylated peptide of E-cadherin (residues 749–761) was titrated at a molar concentration of 575 μM against 75 μM of $HYB^{\Delta C}$ in a VP-ITC microcalorimeter (Microcal, Northhampton, UK) at 293 K. 30 titrations were carried out using 10- μl injections of the peptide into the sample cell containing $HYB^{\Delta C}$ and the data were

analysed with a single-site binding model using the Origin software package v7.0 supplied by Microcal. All measurements were repeated thrice.

### 3.2.7 Analytical Ultracentrifugation (AUC)

The oligomeric states of the $HYB^{\Delta C}$ in the absence and presence of pTyr-peptide was investigated by monitoring their sedimentation properties in the AUC sedimentation velocity experiments. 400 µl of samples at $A_{280\,nm}$ of 1 in 50mM Sodium Phosphate pH 6.5 and 5 mM DTT were used. Sedimentation velocity profiles were collected by monitoring the absorbance at 280 nm. The samples were centrifuged at 40,000 rpm at 25 °C in a Beckman ProteomeLab XL-I centrifuge fitted with a four-hole AN-60 Ti rotor and double-sector aluminum centerpieces and equipped with absorbance optics. The scans were analyzed using the Sedfit program (Brown and Schuck, 2006).

## 3.3 Results

### 3.3.1 $HYB^{\Delta C}$ is a monomer in solution

We previously demonstrated that the HYB domain consists of an atypical zinc-coordinated, intertwined homodimer formed by an anti-parallel arrangement of two Hakai (aa 106–206) monomers (Mukherjee et al., 2012). The dimerization mainly occurs through the C-terminal region of Hakai (aa 106-206), which harbors key pTyr-interacting residues as well as the atypical zinc-coordination motif formed by two histidine residues (H185 and H190) and one cysteine residue (C172) from one monomer and a second cysteine residue (C166) from the adjacent monomer. In this study, we expressed and purified a C-terminal deletion mutant comprising residues

106-194, herein referred to as HYB$^{\Delta C}$, which retains all of the residues involved in the atypical zinc-coordinated motif and the pTyr interaction. The gel-filtration elution profile showed that HYB$^{\Delta C}$ had an apparent molecular weight equivalent to a monomeric subunit (Figure 3.1). Further, the monomeric nature of HYB$^{\Delta C}$ was verified using AUC analysis, which showed a peak corresponding to an apparent molecular weight of monomeric HYB$^{\Delta C}$ (~10,200 Da) (Figure 3.2). The dynamic light scattering experiments also revealed an apparent molecular weight of monomeric HYB$^{\Delta C}$. The dynamic light scattering experiments also revealed an apparent molecular weight of the monomeric HYB$^{\Delta C}$ (Figure 3.3).

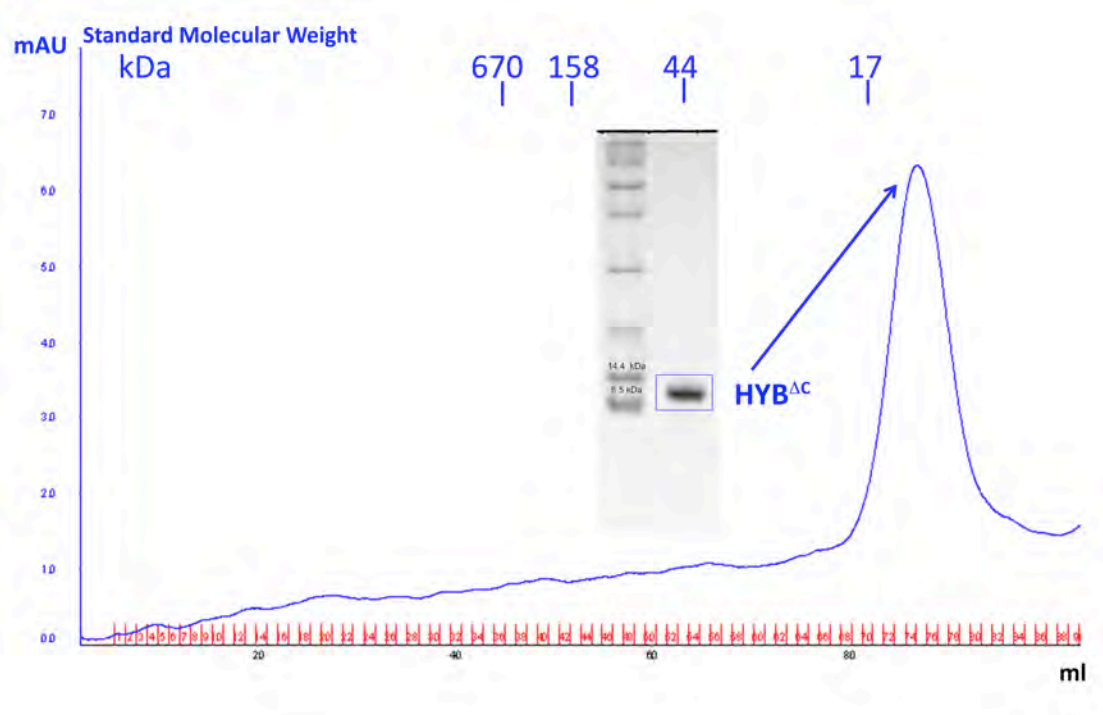



**Figure 3.1** Gel Filtration Profile of HYB$^{\Delta C}$. HYB$^{\Delta C}$ was loaded onto a calibrated Superdex-75 gel filtration chromatography column. The elution profile suggests that HYB$^{\Delta C}$ exists as a monomer. The SDS-page of HYB$^{\Delta C}$ depicts the purity and the molecular weight of the peak fraction (~10kDa).

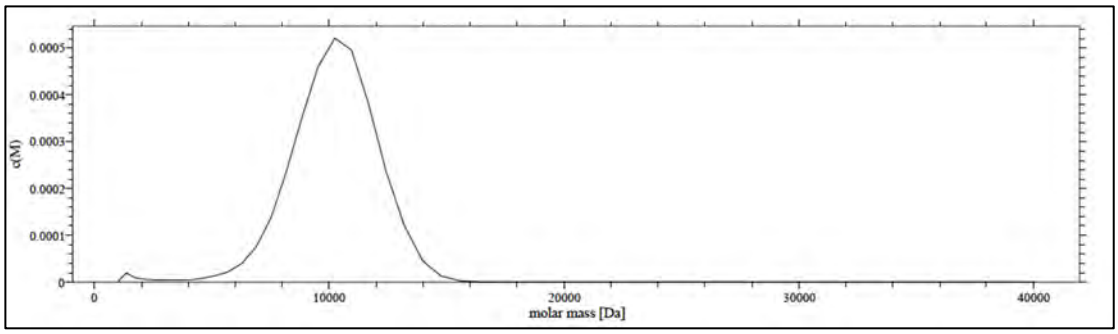



**Figure 3.2** Analytical ultracentrifugation (AUC) analysis of HYB$^{\Delta C}$. The monomeric nature of HYB$^{\Delta C}$ was studied using sedimentation velocity analysis. The sedimentation velocity profiles were collected by monitoring the absorbance at 280 nm as the samples with concentrations at $A_{280\ nm}$ of 1.0 were centrifuged in a four-hole AN-60 Ti rotor at 40,000 rpm at 25 °C. The scans were analyzed using Sedfit program (Brown and Schuck, 2006) to obtain the molecular mass profile of HYB$^{\Delta C}$ which indicates that the protein exists as a monomer in solution with an apparent molecular weight of 10kDa.

| Msr# | Time(s) | Temp(C) | Count Rate | Ampl | Diff Coeff | Radius(nm) | Polyd(nm) | PolydIndx | MW(KDa) | %Mass | Baseline | Sos Error |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 10.0 | 20.0 | 157676 | 0.652 | 1.13e+003 | 1.86 | 0.434 | 0.05 | 14.3 | 100.0 | 1.002 | 4.82 |
| 2 | 20.0 | 20.0 | 151003 | 0.645 | 1.15e+003 | 1.83 | 0.286 | 0.02 | 13.9 | 100.0 | 1.000 | 3.43 |
| 3 | 30.0 | 20.0 | 148806 | 0.742 | 1.13e+003 | 1.85 | 0.270 | 0.02 | 14.2 | 100.0 | 1.000 | 5.50 |
| 4 | 40.0 | 20.0 | 145603 | 0.748 | 1.11e+003 | 1.89 | 0.424 | 0.05 | 14.8 | 100.0 | 1.000 | 5.01 |
| 5 | 50.0 | 19.9 | 147052 | 0.740 | 1.09e+003 | 1.92 | 0.356 | 0.03 | 15.5 | 100.0 | 1.000 | 2.91 |
| 6 | 60.0 | 19.9 | 150311 | 0.733 | 1.08e+003 | 1.93 | 0.271 | 0.02 | 15.7 | 100.0 | 1.000 | 5.65 |
| 7 | 70.0 | 19.9 | 148199 | 0.726 | 1.08e+003 | 1.94 | 0.282 | 0.02 | 16.0 | 100.0 | 1.000 | 4.24 |
| 8 | 80.0 | 20.0 | 150363 | 0.717 | 1.06e+003 | 1.97 | 0.195 | 0.01 | 16.5 | 100.0 | 1.000 | 3.61 |
| 9 | 90.0 | 20.0 | 149415 | 0.715 | 1.08e+003 | 1.95 | 0.292 | 0.02 | 16.0 | 100.0 | 1.000 | 3.34 |
| 10 | 100.0 | 20.0 | 152284 | 0.724 | 1.06e+003 | 1.98 | 0.313 | 0.03 | 16.6 | 100.0 | 1.000 | 2.65 |
| 11 | 110.0 | 20.0 | 155130 | 0.729 | 1.04e+003 | 2.02 | 0.282 | 0.02 | 17.5 | 100.0 | 1.000 | 4.17 |
| 12 | 120.0 | 20.0 | 154283 | 0.726 | 1.07e+003 | 1.96 | 0.304 | 0.02 | 16.3 | 100.0 | 1.000 | 4.50 |
| 13 | 130.0 | 20.0 | 154623 | 0.717 | 1.07e+003 | 1.96 | 0.408 | 0.04 | 16.3 | 100.0 | 1.000 | 3.63 |
| 14 | 140.0 | 20.1 | 154088 | 0.716 | 1.07e+003 | 1.97 | 0.0860 | 0.00 | 16.5 | 100.0 | 1.000 | 3.23 |
| 15 | 150.0 | 20.1 | 153750 | 0.707 | 1.07e+003 | 1.96 | 0.0632 | 0.00 | 16.3 | 100.0 | 1.000 | 2.44 |
| Aves: | | | | | | | | | | | | |
| Mono | | 20.0 | 151505 | 0.716 | 1.09e+003 | 1.93 | 0.284 | 0.02 | 15.8 | 100.0 | 1.000 | 3.94 |
| Bi-1 | | 0.0 | 0 | 0.000 | 0.000 | 0.000 | ---- | ---- | 0.000 | 0.0 | 0.000 | 0.000 |
| Bi-2 | | | | 0.000 | 0.000 | 0.000 | | | 0.000 | 0.0 | | |

**Figure 3.3** DLS studies performed for HYB$^{\Delta C}$. The DLS readout of HYB$^{\Delta C}$ with concentration at $A_{280nm}$ of 1.0 shows an apparent molecular weight close to the monomeric HYB$^{\Delta C}$. The apparent molecular weight is highlighted by a red box.

### 3.3.2 Circular dichroism indicates conformational changes in HYB$^{\Delta C}$ in the presence of substrate

Circular dichroism (CD) analysis revealed significant differences in the secondary structure composition between the monomeric HYB$^{\Delta C}$ and the dimeric HYB domain (Figure 3.4). In the presence of the phosphorylated peptide of E-cadherin,

corresponding to residues 747-760, the CD spectra demonstrated conformational changes to the HYB$^{\Delta C}$ (Figure 3.4). These significant differences in the conformation of HYB$^{\Delta C}$ in the CD spectral analysis motivated us to determine the NMR structure of HYB$^{\Delta C}$ to understand the structural basis of this ligand-associated conformational change.

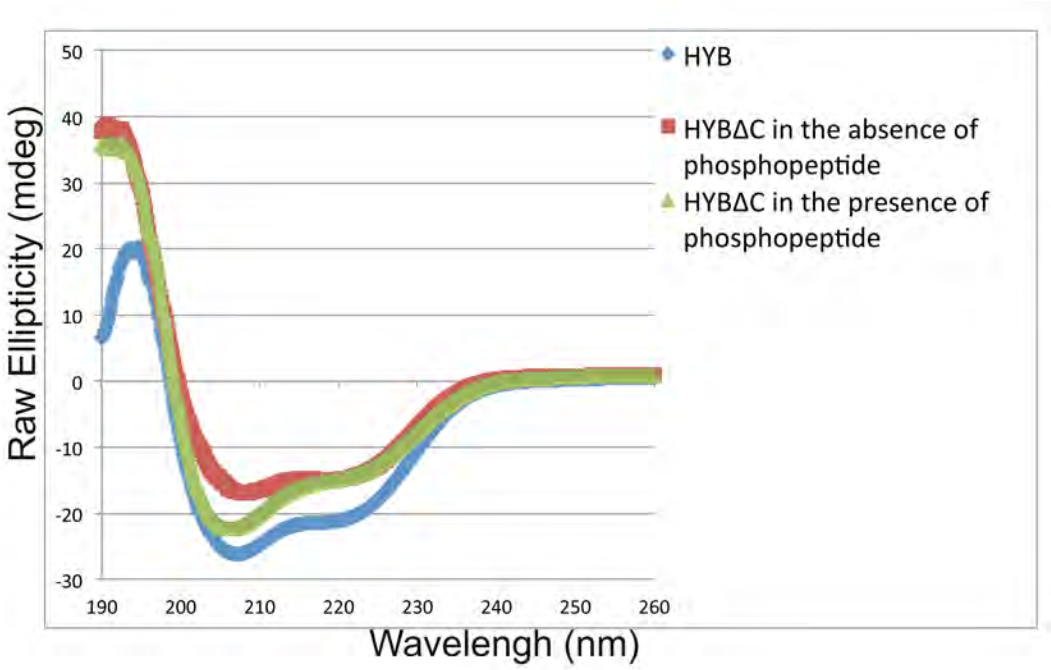



**Figure 3.4** Circular Dichroism analyses revealed differences between HYB$^{\Delta C}$ and HYB. Circular dichroism analysis shows a significant difference in the secondary structure composition between HYB$^{\Delta C}$ and HYB domain. Furthermore, in the presence of the phosphopeptide ligand (E-cadherin$^{747-760}$), the CD spectrum of HYB$^{\Delta C}$ shows marked resemblance with that of the HYB domain, suggesting a conformational change in HYB$^{\Delta C}$ in the presence of ligand.

### 3.3.3 NMR structure of HYB$^{\Delta C}$

In order to gain structural insights into the tertiary fold of HYB$^{\Delta C}$, we solved its solution structure by NMR. The NMR structure of HYB$^{\Delta C}$ was refined to a final RMSD of 1.42 ± 0.44 Å (backbone) (Table 3.1) for the 20 best structures (Figure 3.5 and Figure 3.6). The NMR structure of HYB$^{\Delta C}$ reveals a monomeric fold, which contains a N-terminal RING domain and a C-terminal zinc co-ordination domain

bearing a C2H2-type zinc finger (ZnF) configuration (Figure 3.6 and 3.7). As expected, the solution structure of $HYB^{\Delta C}$ contains three Zinc co-ordination sites (Figure 3.6 and 3.7). The RING domain (residues 106-148) adopts a typical RING architecture containing two β-strands (β1 and β2) and one α-helix (α1), which engage two Zinc ions (Figures 3.6 and 3.7). The C-terminal domain (residues 148-194) lies in close association with the N-terminal RING domain by means of the third β-strand, β3, which forms an anti-parallel β-sheet arrangement with β1 and β2 of the RING domain (Figure 3.6 and 3.7). The C-terminal domain contains the third Zinc co-ordination site of $HYB^{\Delta C}$, which is formed by the residues C166, C172, H185 and H190, resembling a C2H2-like ZnF (Figure 3.7), with the two cysteines (C166 and C172) forming a hairpin-loop and the two histidines (H185 and H190) lying on the α-helix (α2).

The overall structure of $HYB^{\Delta C}$ was compared with other proteins present in the Protein Data Bank (PDB) using the DALI server (http://ekhidna.biocenter.helsinki.fi/dali_server) (Holm et al., 2008). The DALI server showed around 230 structural homologues, all of which contain a RING domain fold like $HYB^{\Delta C}$, but no structural resemblance beyond the RING domain (Table 3.2). In order to identify if any protein has a fold similar to the C-terminal region of $HYB^{\Delta C}$ that contains the C2H2-ZnF, a DALI search was performed for the localized fold comprising of the amino-acid residues 159-194 of $HYB^{\Delta C}$. The DALI results showed that there are only 4 proteins, which contain a fold similar to the C2H2-ZnF of $HYB^{\Delta C}$ (Table 3.3). Interestingly, none of these proteins contain a RING domain. Thus, the structural analysis using DALI clearly indicates that the combination of the

N-terminal RING domain and the C-terminal region containing the C2H2-ZnF produces a novel fold in $HYB^{\Delta C}$.

| **Parameters** | |
|---|---|
| **All NOE distance restraints**[a] | 1014 |
| Intra-residue | 240 |
| Sequential (\|i–j\| = 1) | 333 |
| Medium-range ( 1 <\|i–j\|≤5) | 197 |
| Long-range (\|i–j\|≥5) | 244 |
| **Hydrogen bonds restraints** | 21 |
| **Dihedral angle restraints(φ, ψ )**[b] | 110 |
| **Deviations from idealized covalent geometry**[c] | |
| RMSD of bond lengths (Å) | 0.0026±0.0026 |
| RMSD of bond angles (°) | 0.3844±0.0224 |
| RMSD of improper angles (°) | 0.3272±0.0182 |
| **Deviations from experimental restraints** | |
| RMSD of distance restraints (Å) | 0.0472±0.0021 |
| RMSD of dihedral angle restraints (°) | 0.2363±0.048 |
| **Ramachandran plot analysis (%)**[d] | |
| Residues in allowed region | 94.4% |
| Residues in generally allowed regions | 5.1% |
| Residues in disallows disallowed regions | 0.5% |
| **Average RMSD from mean structure (Å)**[e] | |
| All residues | 1.42±0.44 |
| Regular 2° structure region | 0.76±0.33 |

**Table 3.1** NMR data and structure determination details for $HYB^{\Delta C}$

[a] The distance restraints were obtained by classifying the NOE cross peaks into three categories: strong (1.8–2.9 Å), medium (1.8–3.5 Å), and weak (1.8–5.0 Å). Solution structures of Hakai was calculated using the torsion angle dynamics simulated annealing method within XPLOR-NIH(Schwieters et al., 2003).

[b] Dihedral angles of backbone ϕ and ψ were predicted by TALOS (Cornilescu et al., 1999) using the chemical shifts of Cα, Cβ, Hα, N, and HN.

[c] Twenty lowest-energy conformers with no NOE violations greater than 0.3 Å and no torsion angle violations greater than 3° were selected from 100 conformers to represent the NMR ensembles.

[d] Calculated with PROCHECK-NMR (Laskowski et al., 1996).

[e] Calculated with MOLMOL(Koradi et al., 1996); averages are over backbone atoms.

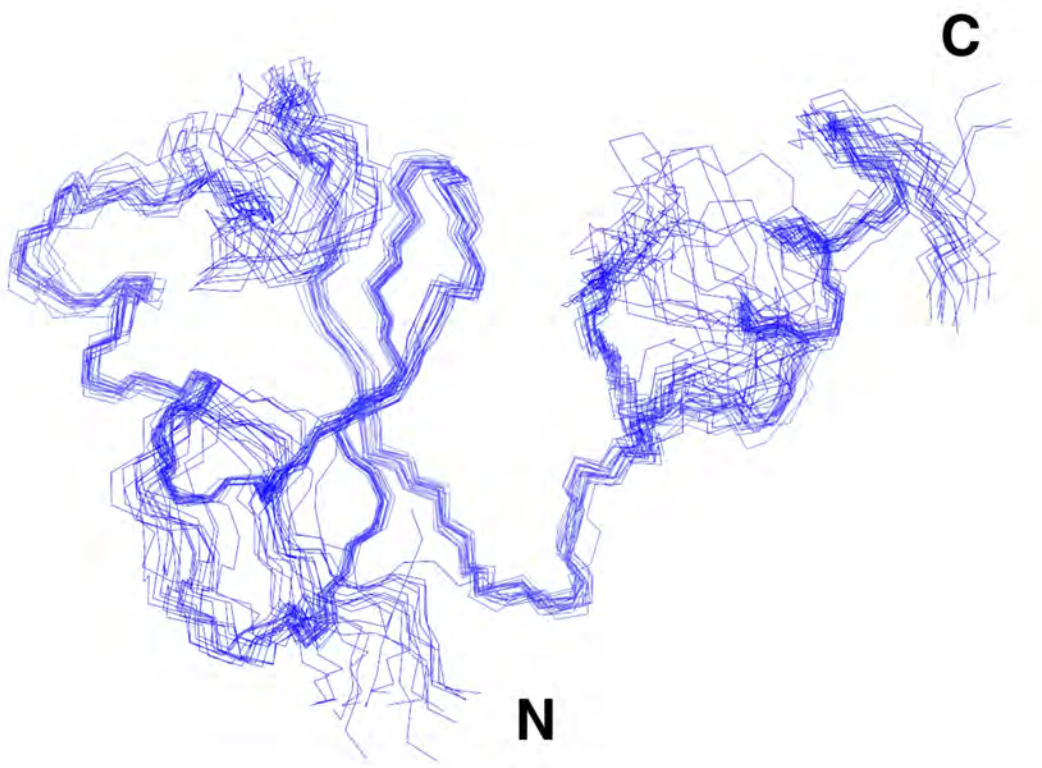



**Figure 3.5** 3D NMR structure of $HYB^{\Delta C}$. Superposition of 20 energy-minimized conformers representing the 3D NMR structure of $HYB^{\Delta C}$

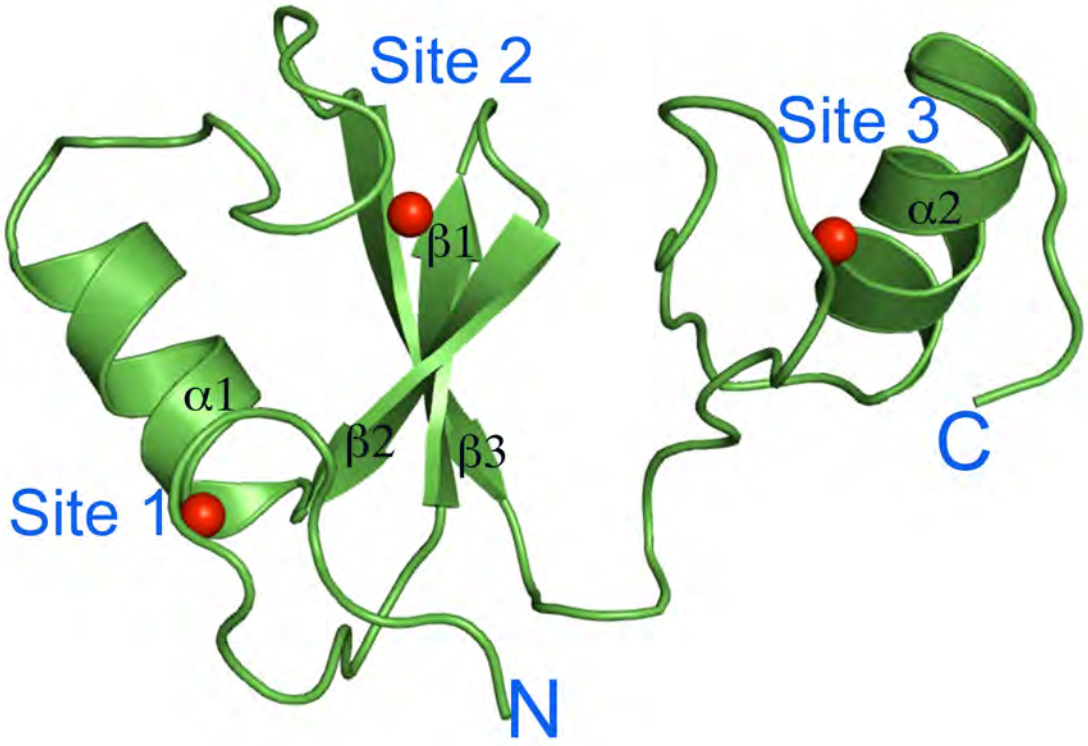



**Figure 3.6** Ribbon diagram of the lowest energy conformer representing the 3D NMR structure of $HYB^{\Delta C}$. Each monomer contains three zinc coordination sites- Site 1, Site 2 and Site 3.The zinc ions are shown as red spheres.

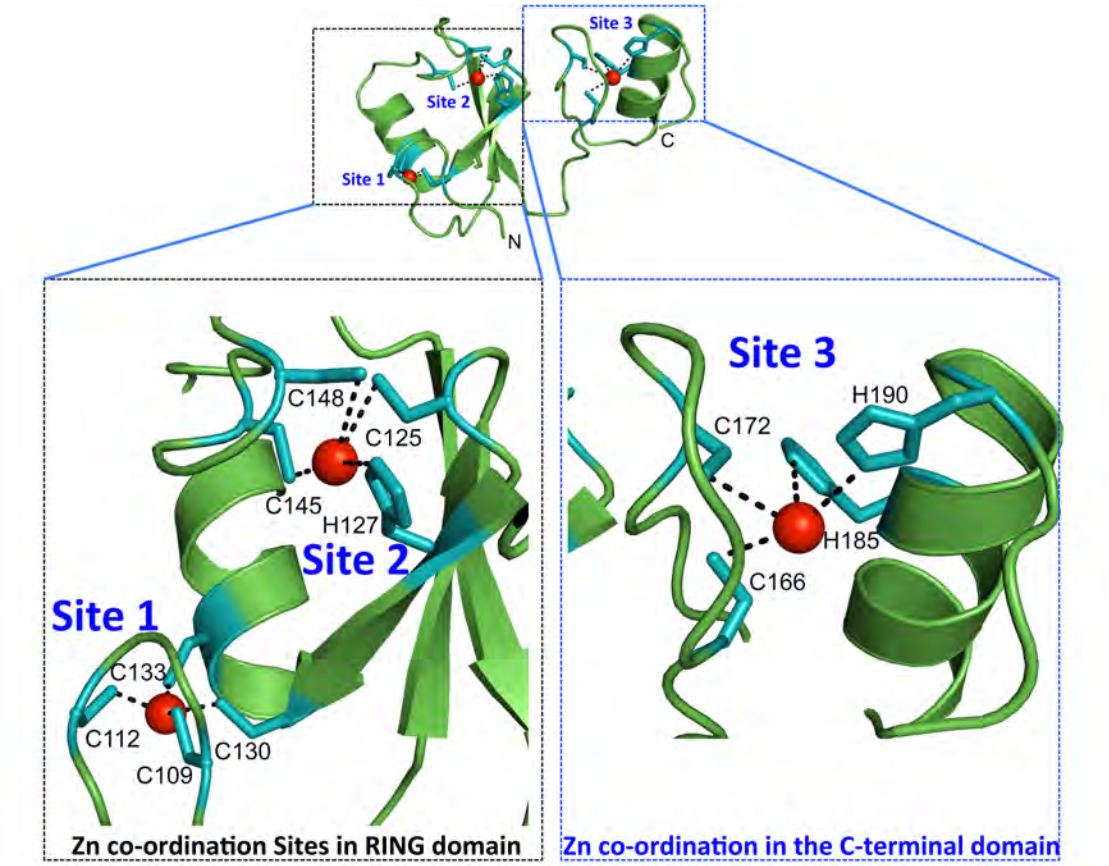



**Figure 3.7** Metal co-ordinations in the NMR structure of $HYB^{\Delta C}$. The coordination of zinc ions (red spheres) in the RING domain and the C-terminal domain of $HYB^{\Delta C}$ are shown.

| DALI Matches | PDB code | Z score | RMSD [Å] | Aligned length | Sequence identity [%] | Protein name |
|---|---|---|---|---|---|---|
| 1 | 3vk6 | 9.8 | 4.6 | 69 | 100 | HYB DOMAIN |
| 2 | 2ckl | 7.6 | 2.1 | 57 | 26 | POLYCOMB GROUP RING FINGER PROTEIN 4 |
| 3 | 2y1n | 7.4 | 2.1 | 56 | 20 | E3 UBIQUITIN-PROTEIN LIGASE |
| 4 | 3rpg | 7.2 | 2.0 | 56 | 27 | UBIQUITIN-CONJUGATING ENZYME E2 D3 |
| 5 | 2oxq | 7.0 | 2.3 | 56 | 5 | UBIQUITIN-CONJUGATING ENZYME E2D 1 |
| 6 | 4epo | 6.9 | 2.0 | 53 | 21 | UBIQUITIN-CONJUGATING ENZYME E2 VARIANT 2 |
| 7 | 2vje | 6.9 | 1.8 | 53 | 26 | E3 UBIQUITIN-PROTEIN LIGASE MDM2 |
| 8 | 2bay | 6.9 | 2.1 | 54 | 13 | PRE-MRNA SPLICING FACTOR PRP19 |
| 9 | 2h0d | 6.8 | 2.0 | 56 | 27 | B LYMPHOMA MO-MLV INSERTION REGION |
| 10 | 2f42 | 6.8 | 2.2 | 56 | 5 | STIP1 HOMOLOGY AND U-BOX CONTAINING PROTEIN 1 |
| 11 | 2c2v | 6.8 | 2.3 | 56 | 5 | CARBOXY TERMINUS OF HSP70-INTERACTING PROTEIN |
| 12 | 3l1x | 6.7 | 2.3 | 55 | 7 | UBIQUITIN CONJUGATION FACTOR E4 B |
| 13 | 4ayc | 6.6 | 2.1 | 55 | 20 | UBIQUITIN-PROTEIN LIGASE RNF8 |
| 14 | 2xeu | 6.5 | 1.8 | 52 | 25 | RING FINGER PROTEIN 4 |
| 15 | 3m63 | 6.4 | 2.4 | 56 | 13 | UBIQUITIN CONJUGATION FACTOR E4 |
| --- | --- | --- | --- | --- | --- | --- |
| 234 | 3lju | 2.1 | 3.2 | 47 | 17 | ARF-GAP WITH DUAL PH DOMAIN-CONTAINING PROTEIN 1 |
| 235 | 3fm8 | 2.1 | 2.7 | 46 | 17 | KINESIN-LIKE PROTEIN KIF13B |
| 236 | 2p57 | 2.0 | 4.3 | 55 | 18 | GTPASE-ACTIVATING PROTEIN ZNF289 |

**Table 3.2** Structural comparison of $HYB^{\Delta C}$ domain with structural homologs. Representative structural matches of the $HYB^{\Delta C}$ searched in the PDB using DALI server (Holm et al., 2008).

| DALI Matches | PDB code | Z score | RMSD [Å] | Aligned length | Sequence identity [%] | Protein name |
|---|---|---|---|---|---|---|
| 1 | 2elx | 2.4 | 1.9 | 27 | 30 | ZINC FINGER PROTEIN 406 |
| 2 | 1sp2 | 2.3 | 1.2 | 26 | 27 | SP1F2 |
| 3 | 3mjh | 2.1 | 1.2 | 25 | 28 | RAS-RELATED PROTEIN RAB-5A |
| 4 | 1zfd | 2.1 | 1.4 | 26 | 19 | SWI5 |

**Table 3.3** Structural comparison of HYB$^{\Delta C}$(amino acid residues 159-194) domain with structural homologs. Representative structural matches of the HYB$^{\Delta C}$ (amino acid residues 159-194) searched in the PDB using DALI server (Holm et al., 2008).

### 3.3.4 Comparison of the structure of HYB$^{\Delta C}$ with HYB domain

In order to analyze the structural differences between monomeric HYB$^{\Delta C}$ with HYB domain, HYB$^{\Delta C}$ was superposed onto one of the Hakai (aa 106-206) monomeric counterparts present in the HYB domain. The structural segment corresponding to residues 106-166 of HYB$^{\Delta C}$ and a monomer equivalent of the HYB domain fold in a similar manner with a RMSD of 0.9 Å (Figure 3.8C). This implies that the RING domain (residues 106-148) and linker sequence (residues 149-166) fold exactly in a similar manner in both HYB$^{\Delta C}$ and the monomer equivalent in the HYB domain. The structural differences arise beyond Cys166, which comprises the C–terminal zinc co-ordination site (Site 3) and the terminal α-helix (α2) (Figure 3.8). The C-terminal helix α2 of HYB$^{\Delta C}$ flips by 180° and is separated from the α2 of HYB monomeric counterpart by 25 Å distance (Figure 3.8C and 3.8D). Furthermore, the comparison of HYB$^{\Delta C}$ with the dimeric HYB domain clearly reveals a structural rearrangement, which arises due to systematic reshuffling among the zinc-coordination residues to engage the zinc ion in the C-terminal region of the monomeric and dimeric conformers, respectively (Figure 3.9). In HYB$^{\Delta C}$, the four zinc-coordination residues

forms a C2H2-ZnF, with all residues situated on the same monomeric chain, which is distinct from the atypical zinc-coordination motif present in HYB domain that is formed by sharing of the zinc ions between two monomers (Figure 3.9). Thus, the NMR structure provides the basis for the existence of $HYB^{\Delta C}$ as a monomer in solution, bearing a different fold, with the reason being the replacement of the atypical zinc-binding motif, which is critical for the dimeric fold of HYB, by a C2H2-type ZnF. The overall topology of $HYB^{\Delta C}$ and the HYB domain is compared in Figure 3.10.

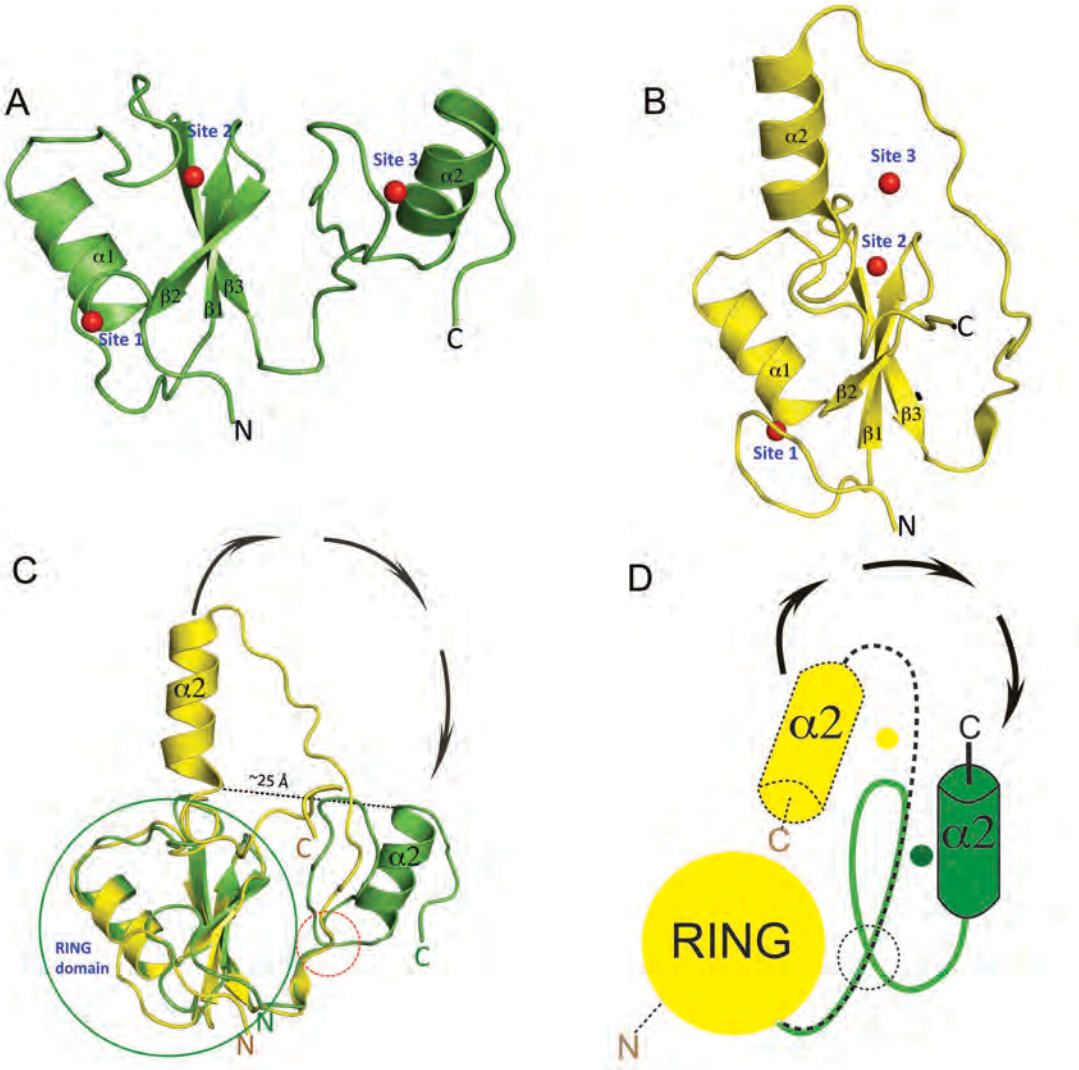



**Figure 3.8** Comparison of the monomeric $HYB^{\Delta C}$ (shown in green) with the Hakai (aa 106-206) monomeric counterpart (shown in yellow) of the dimeric HYB domain.

(A) and (B) Ribbon representations of HYB$^{\Delta C}$ and Hakai (aa 106-206) monomeric counterpart (pdb code: 3vk6) of HYB domain shown respectively, in the same orientation. (C) Superposition of HYB$^{\Delta C}$ on Hakai (aa 106-206) (pdb code: 3vk6) reveals that the stretch of residues from 106-166, which contains the RING domain and a part of the C-terminal domain in both structures overlap with a rmsd of 0.9 Å. The differences arise after C166 highlighted by red dotted circle. Beyond this point, the folds of HYB$^{\Delta C}$ and HYB monomeric counterpart is completely different, with the helix α2 flipping by 180° and separated by around 25 Å with respect to each-other. (D) The conformational changes in the structures of HYB$^{\Delta C}$ and Hakai (aa 106-206) of the HYB domain shown in (C) is schematically represented.

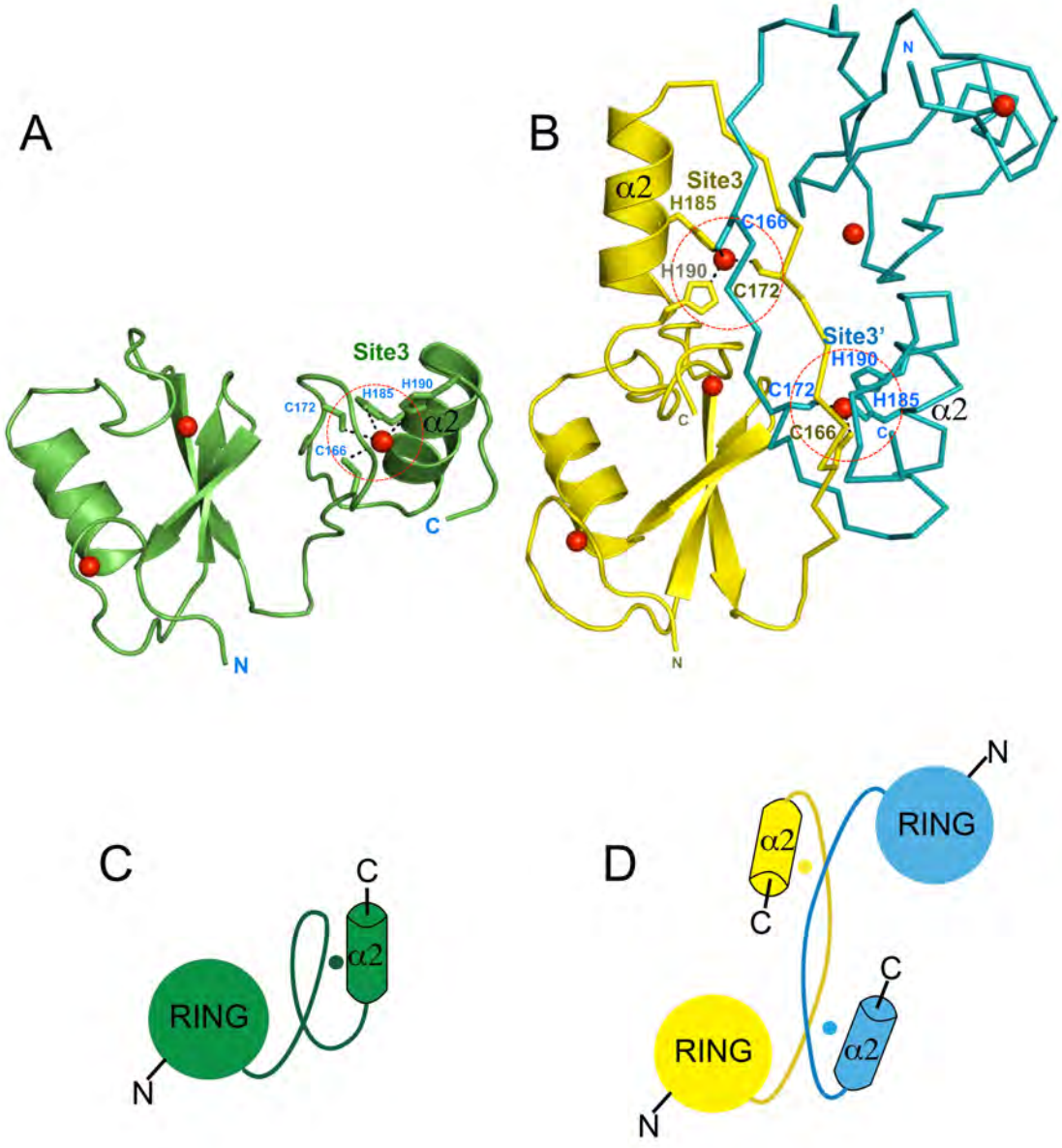



**Figure 3.9** Comparison of the monomeric HYB$^{\Delta C}$ with the HYB domain (pdb code: 3vk6) that contains a dimeric fold containing paired Hakai (aa 106-206) monomers. The monomeric fold of HYB$^{\Delta C}$ (shown in green) contains a C2H2-ZnF in the C-terminal domain, highlighted by dotted circles (A) in which all the four zinc

coordinating residues are situated in the same monomeric chain. On the other hand, the dimeric fold of HYB domain is formed by the sharing of two zinc ions between the Hakai (aa 106-206) monomers, shown by two dotted circles in (B), such that three of the four zinc-binding residues come from one monomer while the fourth residue comes from the adjacent monomer. The two monomers of Hakai (aa 106-206) in (B) are colored as yellow and cyan, with one monomer shown in ribbon representation while the other monomer is shown as Cα trace for clarity. The zinc-coordinating side-chains are shown as sticks, and color-coded according to the color of the monomer on which they lie. (C) and (D) The schematic representation of the difference in the monomeric fold of HYB$^{\Delta C}$ and the dimeric fold of HYB domain containing paired Hakai (aa 106-206) monomers, respectively.

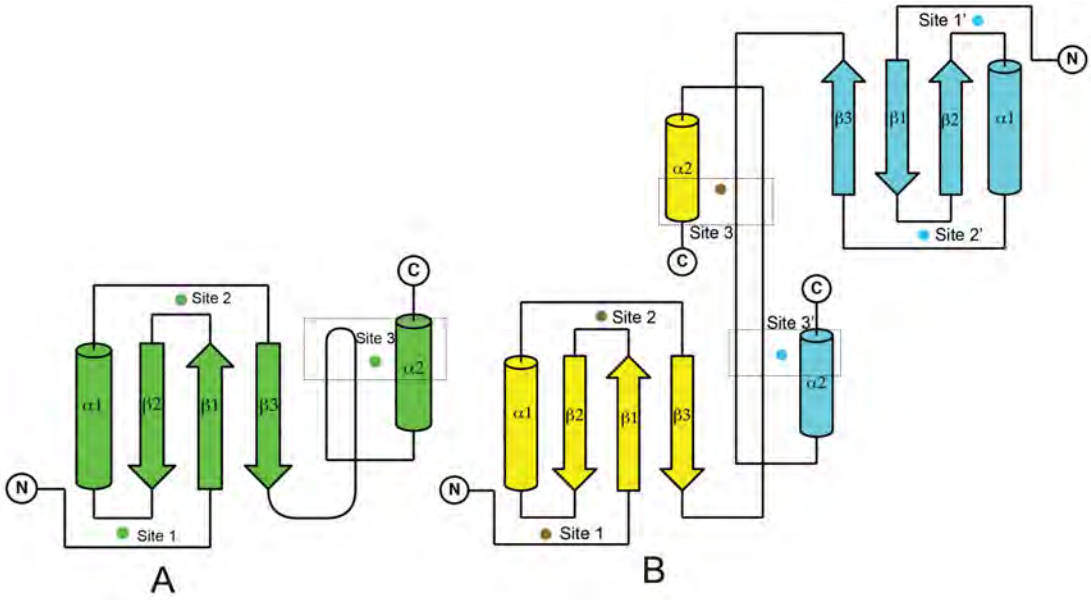
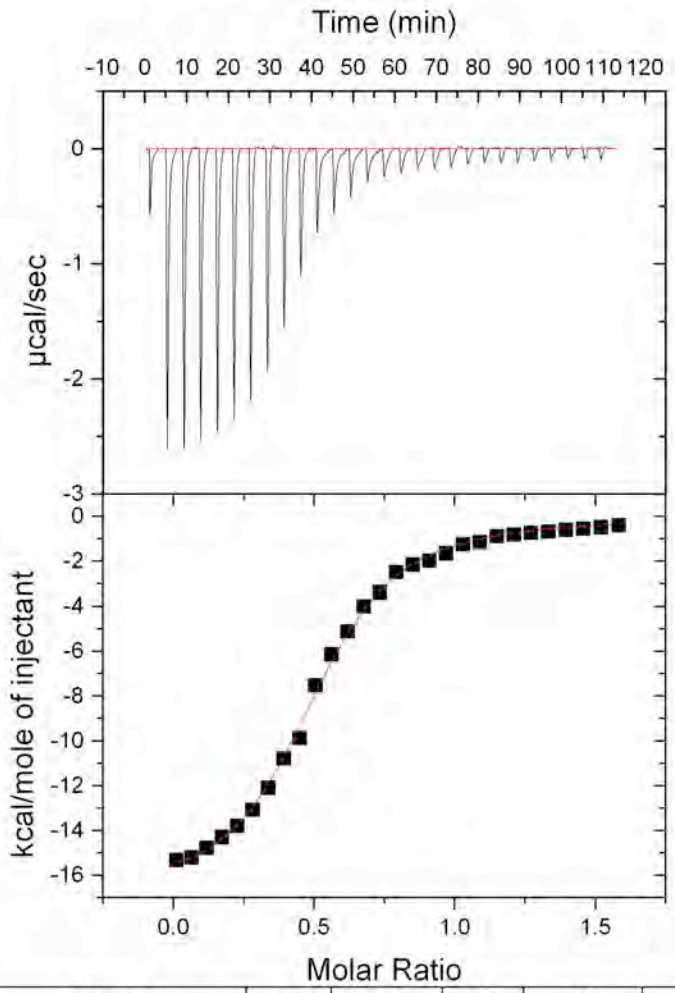



**Figure 3.10** Topology of HYB$^{\Delta C}$ (A) is compared with that of HYB domain (pdb code: 3vk6) (B). The zinc ions are shown as spheres (color coded according to the monomer colors) and the three-zinc coordination sites are indicated for each monomer in both A and B. Each monomer consists of three Zinc coordination sites. In HYB$^{\Delta C}$, four zinc-coordination residues, all situated on the same monomer form the Site 3; while in HYB domain, Site 3 and Site 3'are shared between two monomers, shown in dotted boxes.

### 3.3.5 HYB$^{\Delta C}$ binds phospho-E-cadherin$^{747-760}$ as a dimer

We next questioned if the monomeric HYB$^{\Delta C}$ can still engage the pTyr ligand. Our earlier studies showed HYB as a dimeric fold interacts with pTyr ligands (Mukherjee et al., 2012). Hence, we conducted isothermal titration calorimetry (ITC) experiments to investigate the binding property of HYB$^{\Delta C}$ with the peptide ligand corresponding to

our previously established pTyr motif of E-cadherin, which contains the E-cadherin residues (747-760) (Mukherjee et al., 2012). The ITC experiment revealed that $HYB^{\Delta C}$ binds with the pTyr ligand of E-cadherin with a binding stoichiometry of 2:1 [$HYB^{\Delta C}$: pTyr ligand] and an affinity of 3.7μM $K_d$ (Figure 3.11). This indicates that phospho-E-cadherin$^{747-760}$ ligand interacts with a dimeric form of $HYB^{\Delta C}$.




| Peptide | Sequence | N | Ka ($x10^6$ $M^{-1}$) | Kd (μM) | ΔH (kcal/mol) | TΔS (kcal/mol) | ΔG (kcal/mol) |
|---|---|---|---|---|---|---|---|
| E-cad$^{747-760}$ | DTRDNVYpYYDEEG | 0.511 ± 0.004 | 0.2706 ± 0.014 | 3.7 | -16.94 ± 0.165 | -9.358 | -7.582 |

**Figure 3.11** Binding studies of $HYB^{\Delta C}$ with phospho-E-cadherin$^{747-760}$ using ITC. The tyrosine phosphorylated E-cadherin$^{747-760}$ was titrated against $HYB^{\Delta C}$ using ITC. The top panels show the heat release profiles after baseline correction and the lower panels indicate the binding isotherms for the interactions. The dissociation constant ($K_d$) and binding stoichiometry (N) is shown in the table.

To further probe the dimeric state of $HYB^{\Delta C}$ in the presence of the phospho-E-cadherin$^{747-760}$ ligand, $HYB^{\Delta C}$ was separated on a calibrated gel-filtration column in the presence and absence of the ligand. The comparison of the gel-filtration elution profiles in the presence and absence of the phospho-E-cadherin$^{747-760}$ ligand show that $HYB^{\Delta C}$ elutes as a dimer in the presence of the ligand and as monomer in its *apo* form (Figure 3.12).

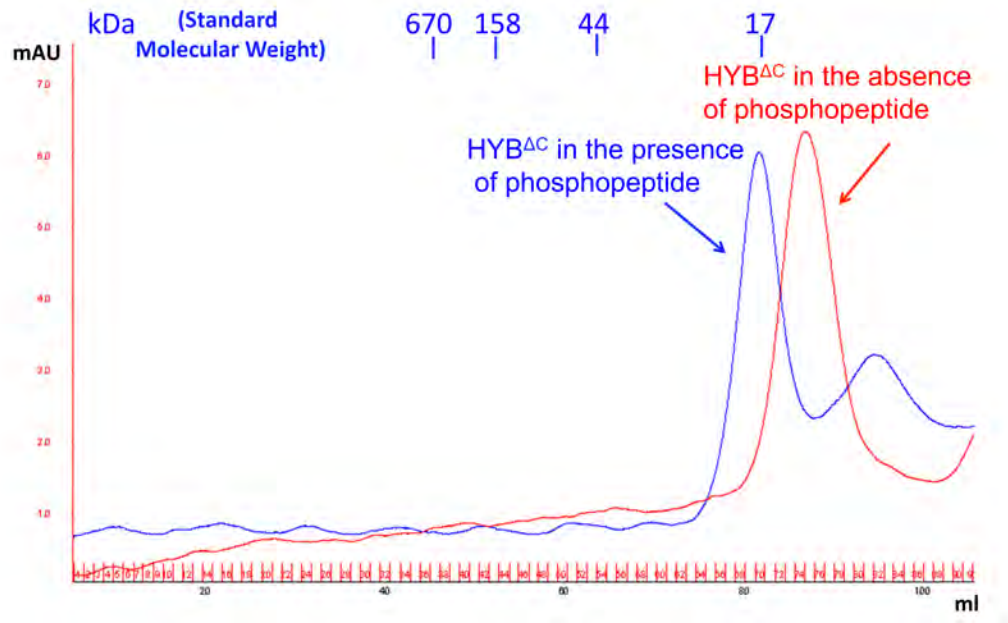



**Figure 3.12** Comparison of gel filtration profiles of $HYB^{\Delta C}$ in the presence and absence of pTyr ligand (phospho-E-cadherin$^{747-760}$). $HYB^{\Delta C}$ was loaded onto a calibrated Superdex-75 gel filtration chromatography column in the presence and absence of the pTyr ligand. The elution profile suggests that $HYB^{\Delta C}$ exists as a dimer in the presence of ligand whereas; it exists as a monomer in the absence of the ligand.

In order to further verify the ligand-induced dimerization of $HYB^{\Delta C}$, sedimentation velocity AUC experiments were conducted with the samples containing $HYB^{\Delta C}$ in the presence of phospho-E-cadherin$^{747-760}$ peptide. From analysis of the AUC data using Sedfit (Brown and Schuck, 2006), the apparent molecular weight of $HYB^{\Delta C}$ in the presence of ligand is 20000Da (Figure 3.13), which is equivalent to the twice the

monomeric molecular weight of HYB$^{\Delta C}$ determined by AUC in the absence of the ligand (Figure 3.2). Additionally, the formation of HYB$^{\Delta C}$ dimer in the presence of ligand is also supported by the results obtained through dynamic light scattering, which show an apparent molecular weight of corresponding to the dimer (Figure 3.14). The combined results suggest that HYB$^{\Delta C}$ switches from a monomeric conformation to a dimeric form in the presence of the pTyr-containing ligand.

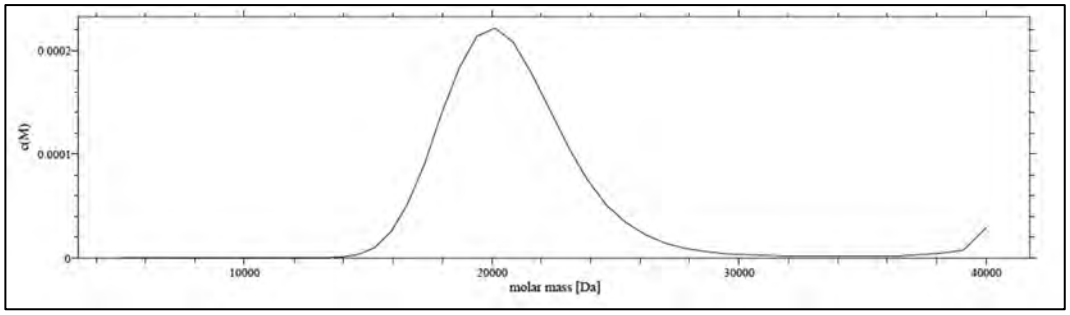



**Figure 3.13** AUC analysis of HYB$^{\Delta C}$ in the presence of phospho-E-cadherin$^{747\text{-}760}$ ligand. The ligand-induced dimerization of HYB$^{\Delta C}$ was studied using by sedimentation velocity analysis. The sedimentation velocity profiles were collected by monitoring the absorbance at 280 nm as the samples with concentration at $A_{280\ nm}$ of 1.0 were centrifuged in a four-hole AN-60 Ti rotor at 40,000 rpm at 25 °C. The scans were analyzed using Sedfit program (Brown and Schuck, 2006) to obtain the molecular mass profile of HYB$^{\Delta C}$ in the presence of the ligand. The results clearly indicate that the protein exists as a dimer in the presence of the ligand with an apparent molecular weight of 20000Da.

| Msr# | Time(s) | Temp(C) | Count Rate | Ampl | Diff Coeff | Radius(nm) | Polyd(nm) | PolydIndx | MW(KDa) | %Mass | Baseline | Sos Error |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 10.0 | 20.1 | 152637 | 0.748 | 1.13e+003 | 1.91 | 0.603 | 0.10 | 15.2 | 100.0 | 1.000 | 7.37 |
| 2 | 20.0 | 20.1 | 150293 | 0.711 | 1.06e+003 | 2.02 | 0.262 | 0.02 | 17.4 | 100.0 | 1.000 | 5.18 |
| 3 | 30.0 | 20.0 | 148872 | 0.706 | 1.04e+003 | 2.06 | 0.261 | 0.02 | 18.3 | 100.0 | 1.000 | 5.82 |
| 4 | 40.0 | 20.0 | 153894 | 0.697 | 1.01e+003 | 2.11 | 0.593 | 0.08 | 19.3 | 100.0 | 1.000 | 4.91 |
| 5 | 50.0 | 20.0 | 152497 | 0.732 | 980. | 2.18 | 0.539 | 0.06 | 21.0 | 100.0 | 1.001 | 8.02 |
| 6 | 60.0 | 20.0 | 157283 | 0.763 | 959. | 2.23 | 0.322 | 0.02 | 22.0 | 100.0 | 1.000 | 8.05 |
| 7 | 70.0 | 20.0 | 155355 | 0.758 | 963. | 2.22 | 0.575 | 0.07 | 21.8 | 100.0 | 1.000 | 10.5 |
| 8 | 80.0 | 20.0 | 155485 | 0.759 | 963. | 2.22 | 0.619 | 0.08 | 21.8 | 100.0 | 1.000 | 8.16 |
| 9 | 90.0 | 20.0 | 158311 | 0.747 | 949. | 2.26 | 0.657 | 0.08 | 22.6 | 100.0 | 1.000 | 6.69 |
| 10 | 100.0 | 20.0 | 155319 | 0.753 | 954. | 2.24 | 0.558 | 0.06 | 22.3 | 100.0 | 1.000 | 7.22 |
| 11 | 110.0 | 20.0 | 154728 | 0.744 | 956. | 2.24 | 0.386 | 0.03 | 22.2 | 100.0 | 1.000 | 8.43 |
| 12 | 120.0 | 20.0 | 154453 | 0.750 | 942. | 2.27 | 0.279 | 0.02 | 22.9 | 100.0 | 1.000 | 5.42 |
| 13 | 130.0 | 20.0 | 154853 | 0.741 | 964. | 2.22 | 0.776 | 0.12 | 21.7 | 100.0 | 1.000 | 4.87 |
| 14 | 140.0 | 20.0 | 154574 | 0.739 | 953. | 2.25 | 0.642 | 0.08 | 22.3 | 100.0 | 1.000 | 7.10 |
| 15 | 150.0 | 20.0 | 155670 | 0.747 | 942. | 2.27 | 0.322 | 0.02 | 23.0 | 100.0 | 1.000 | 5.26 |
| 16 | 160.0 | 20.0 | 154671 | 0.739 | 960. | 2.23 | 0.624 | 0.08 | 22.0 | 100.0 | 1.000 | 9.37 |
| 17 | 170.0 | 20.0 | 154982 | 0.753 | 952. | 2.25 | 0.770 | 0.12 | 22.4 | 100.0 | 1.000 | 7.25 |
| Aves: | | | | | | | | | | | | |
| Mono | | 20.0 | 154345 | 0.740 | 981. | 2.19 | 0.517 | 0.06 | 21.1 | 100.0 | 1.000 | 7.04 |
| Bi-1 | | 0.0 | 0 | 0.000 | 0.000 | 0.000 | ---- | ---- | 0.000 | 0.0 | 0.000 | 0.000 |
| Bi-2 | | | | 0.000 | 0.000 | 0.000 | | | 0.000 | 0.0 | | |

**Figure 3.14** DLS studies performed for HYB$^{\Delta C}$ in the presence of phospho-E-cadherin$^{747\text{-}760}$ ligand. DLS was carried out with HYB$^{\Delta C}$ protein samples with concentration at $A_{280nm}$ of 1.0. The DLS readout indicates that HYB$^{\Delta C}$ exists as a dimer in solution in the presence of ligand.

### 3.4 Discussion

Phosphotyrosine binding domains are critical, modular components that bind to phosphorylated tyrosine residues in acceptor proteins to create multi-protein complexes and regulate several intracellular signalling pathways (Evans et al., 2012; Pawson and Nash, 2003; Yaffe, 2002). Dysregulation of these pathways is often associated with oncogenic transformation (Pawson, 2002), rendering pTyr-binding domains as attractive targets for directed therapies (Burke et al., 2001; Machida and Mayer, 2005b; Sawyer et al., 2002). So far, all of the major pTyr-binding domains, including SH2 and PTB domains, tend to be monomeric (Liu et al., 2006; Stein et al., 2003), barring the few exceptions where they function as homodimers. These exceptions include the SH2 domains of Grb10, Grb14, Grb7, APS, and SH2-B (Depetris et al., 2005; Hu and Hubbard, 2005; Liu et al., 2006; Stein et al., 2003). In addition, in STAT proteins, tyrosine phosphorylation-mediated dimerization has been reported to occur via SH2 domains (Darnell, 1997; Soler-Lopez et al., 2004; Wenta et al., 2008). The recently identified pTyr-binding fold in Hakai, HYB, bears a novel dimeric fold consisting of two atypical ZnFs shared between the paired Hakai (aa 106-206) monomers that exist in an intertwined configuration across the flexible C-terminal regions (Mukherjee et al., 2012). In this study, we showed that the deletion of flexible C-terminal residues of HYB induces a monomeric fold in solution. A similar observation was recently reported in the C-terminal domain of SARS-CoV main protease ($M^{pro\text{-}C}$), where the truncation of the disordered C-terminal helix results in transition from dimer to monomer (Kang et al., 2012). In yet another study, dimer/tetramer equilibrium observed in Escherichia coli DNA mismatch repair (MMR) protein MutS is converted into a monomer/dimer equilibrium upon deletion of the C-terminal 53 amino acids (Manelyte et al., 2006).

In order to understand the structural basis for this monomeric conformation of $HYB^{\Delta C,}$ we determined its solution NMR structure. The NMR structure revealed that the characteristic atypical ZnF, which plays a key role in the dimeric fold of the HYB domain, is replaced by a C2H2-type ZnF in $HYB^{\Delta C}$, formed by two cysteines and two histidines, all situated on the same monomer. This represents a unique monomeric to dimeric switching mechanism.

One of the best studied physiological functions of the HYB domain is its interaction with and regulation of the tyrosine phosphorylated E-cadherin at cell junctions (Fujita et al., 2002; Ishiyama and Ikura, 2012; Ishiyama et al., 2010; Mukherjee et al., 2012). As such, we selected a tyrosine phosphorylated E-cadherin$^{747\text{-}760}$ peptide as a model substrate to further assess this monomeric conformation of $HYB^{\Delta C}$ and to determine if it is still functional. The ITC interaction studies showed that the ligand binds with the protein only in its dimeric form. This result is consistent with our previous studies that a dimeric HYB fold is necessary to create the pTyr-binding pocket, which was also validated with full-length Hakai in cell-based experiments (Mukherjee et al., 2012). A series of solution studies further validated this switch to a dimeric conformation of $HYB^{\Delta C}$ in the presence of the tyrosine phosphorylated E-cadherin$^{747\text{-}760}$ ligand. Previous studies have demonstrated ligand-mediated dimerization as a novel mechanism for protein-carbohydrate recognition (Flint et al., 2004; Sanchez-Vallet et al., 2013). The present study extends this paradigm to pTyr-signaling and regulation in Hakai.

In conclusion, we have identified a novel monomeric fold of $HYB^{\Delta C}$, which impairs

its pTyr binding property. However, in the presence of the pTyr ligand, the monomeric $HYB^{\Delta C}$ becomes a dimer in order to create the pTyr-binding pocket to engage the substrate. The findings from this study suggest that the dimeric architecture of the HYB domain is necessary to engage the pTyr-ligand, which is in sharp contrast to all the other known pTyr-binding domains, which mainly function as monomers. Selectively targeting the dimeric interface of therapeutically important enzymes has emerged as an attractive method of allosteric inhibition (Andreola, 2009; Huber et al., 2012; Lebon and Ledecq, 2000; McMillan et al., 2000). The importance of the dimeric architecture of HYB in pTyr substrate binding demonstrated by the present study makes it an ideal target for designing selective allosteric inhibitors that abrogates HYB dimerization and act as novel therapeutic interventions against cancer.

# CHAPTER IV

# Conclusions and Future Directions

## 4.1 Conclusions

Prior to our studies, SH2 and PTB domains were the only two major pTyr-binding domains known to exist, apart from the idiosyncratic cases exemplified by the C2 domain of PKCδ and the M2 isoform of Pyruvate Kinase. When SH2 domain was discovered about 25 years ago, only 26 proteins were thought to contain SH2 domain. At present, around 111 human proteins have been found to bear a SH2 domain. Similarly, at the time of discovery of PTB domain in 1994, 6 PTB domain-containing proteins were identified, but over the time, more PTB domains have been discovered and presently around 54 PTB domains have been identified in human genome. After a span of more than 18 years, we have discovered a novel pTyr-binding domain in Hakai, and named it as the HYB domain. In the case of HYB, we have identified that it exists across numerous species, ranging from drosophila to human, with at least 4 unique human proteins sequences known to-date. This makes HYB domain the third major pTyr-binding domain after SH2 and PTB domains, which were so far, the only major pTyr-binding domains known to exist. All the HYB homologous protein sequences identified across the numerous species contain the highly conserved 9 cysteines and 3 histidines that contribute to the metal co-ordination, dimerization and to maintain the fold. Thus all these proteins are expected to bear the unique HYB fold. As was the case with SH2 and PTB domains, with more research, it is expected to discover more proteins, which bear the HYB fold. Moreover, the HYB fold has no resemblance with any of the known 75,000 structures in the protein data bank. The novel features of the Hakai pTyr-binding domain guarantee a novel-binding mode with the substrate proteins with which it binds in a pTyr-dependent manner.

Having discovered a novel pTyr-binding structure in Hakai, we next sought to study its binding mechanism with the pTyr-motif of E-cadherin. One of the novel features of the pTyr-motif of E-cadherin was the presence of three consecutive Tyrosine residues. Earlier studies by Fujita et al. in 2002, suggested that the first two tyrosine residues where important for binding with Hakai in a phosphotyrosine dependent manner. However, there were no comprehensive studies, which evaluated the relative importance of the three-tyrosine residues independently and in combination. We identified that a single phosphotyrosine residue of E-cadherin is most critical and can bind independently of remaining two tyrosines present in the binding motif of E-cadherin. In addition, we show that the acidic residues that flank the central phosphotyrosine are critical for binding with Hakai pTyr-binding domain. The NMR and mutational studies revealed that both the RING domain, the minimum phosphotyrosine binding region, as well as dimerization, seen in the structure of Hakai (aa 106-206) are important for recognizing the pTyr-containing motifs. We demonstrated that in conjunction with the RING domain, dimerization of Hakai through the minimum pTyr-binding region leading to the formation of the atypical ZnF creates a novel HYB domain. Furthermore, using mass spectrometry, we experimentally identified around 30 novel Src substrates that are targeted by the HYB domain. Out of these new HYB binding proteins that we discovered, we chose two candidates, *viz.* Cortactin and DOK1, and further characterized their interaction with HYB domain using biophysical and cell-based methods.

In addition, we show that ZNF645 possesses a HYB domain but demonstrates different target specificities. In an additional study, we also demonstrated that the deletion of the C-terminal residues from the HYB domain results in a monomeric

form. However, this monomeric form switches into a dimer in the presence of phospho-E-cadherin peptide in order to interact with the pTyr- motif. This study further confirms that dimerization is absolutely essential for the pTyr-binding property of the HYB domain.

### 4.2 Future directions

The present study provides key insights into the nature and substrate binding mechanism in the novel pTyr-binding domain present in Hakai. Furthermore, we have experimentally validated the pTyr motif that represents the minimum interacting regions present in E-cadherin, Cortactin and DOK1 and binds with the HYB domain with substantial affinity. This forms the basis for carrying out the co-crystallization studies of HYB domain with the minimum interacting regions of E-cadherin, Cortactin and DOK1. Since the interactions of Hakai with E-cadherin, cortactin and DOK1 are physiologically important and implicated in various cancer and diseased conditions, the novel structural features of co-crystal structures will open new avenues for therapeutic interventions. Based on the co-crystal structures of HYB-substrate complexes, small molecule inhibitors can be designed to selectively break these interactions, which may be used as directed therapies towards specific types of cancer that involve such interactions.

In addition to Cortactin and DOK1, we have also identified around 30 binding partners for HYB domain, many of which represent adaptors, kinases, phosphatases, and regulators of small GTPases, which play crucial roles in cell signaling. In future, we propose to further investigate the binding mechanism of the HYB domain with these important targets.

Furthermore, ZNF645 possesses a HYB domain but demonstrates different target specificities, as it can bind tyrosine phosphorylated E-cadherin, but not Cortactin. Therefore, it would be of great interest to study the structural and mechanistic aspects of the HYB domain of ZNF645, in order to understand the determinants of substrate specificity of the HYB domain present in different proteins.

# References


Abella, V., Valladares, M., Rodriguez, T., Haz, M., Blanco, M., Tarrio, N., Iglesias, P., Aparicio, L.A., and Figueroa, A. (2012). miR-203 regulates cell proliferation through its influence on Hakai expression. PloS one *7*, e52568.

Abram, C.L., and Courtneidge, S.A. (2000). Src family tyrosine kinases and growth factor signaling. Experimental cell research *254*, 1-13.

Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., and Terwilliger, T.C. (2002). PHENIX: building new software for automated crystallographic structure determination. Acta crystallographica Section D, Biological crystallography *58*, 1948-1954.

Anastas, J.N., and Moon, R.T. (2013). WNT signalling pathways as therapeutic targets in cancer. Nature reviews Cancer *13*, 11-26.

Andreola, M.L. (2009). Therapeutic potential of peptide motifs against HIV-1 reverse transcriptase and integrase. Current pharmaceutical design *15*, 2508-2519.

Aparicio, L.A., Valladares, M., Blanco, M., Alonso, G., and Figueroa, A. (2012). Biological influence of Hakai in cancer: a 10-year review. Cancer metastasis reviews *31*, 375-386.

Aravind, L., and Koonin, E.V. (2000). The U box is a modified RING finger - a common domain in ubiquitination. Current biology : CB *10*, R132-134.

Ardley, H.C., and Robinson, P.A. (2005). E3 ubiquitin ligases. Essays in biochemistry *41*, 15-30.

Avizienyte, E., and Frame, M.C. (2005). Src and FAK signalling controls adhesion fate and the epithelial-to-mesenchymal transition. Current opinion in cell biology *17*, 542-547.

Bax, A., and Grzesiek, S. (1993). Methodological Advances in Protein Nmr. Accounts Chem Res *26*, 131-138.

Bellon, S.F., Rodgers, K.K., Schatz, D.G., Coleman, J.E., and Steitz, T.A. (1997). Crystal structure of the RAG1 dimerization domain reveals multiple zinc-binding motifs including a novel zinc binuclear cluster. Nature structural biology *4*, 586-591.

Benes, C.H., Wu, N., Elia, A.E., Dharia, T., Cantley, L.C., and Soltoff, S.P. (2005). The C2 domain of PKCdelta is a phosphotyrosine binding domain. Cell *121*, 271-280.

Bisacchi, D., Zhou, Y., Rosen, B.P., Mukhopadhyay, R., and Bordo, D. (2006). Crystallization and preliminary crystallographic characterization of LmACR2, an arsenate/antimonate reductase from Leishmania major. Acta Crystallogr F *62*, 976-979.

Boggon, T.J., and Eck, M.J. (2004). Structure and regulation of Src family kinases. Oncogene *23*, 7918-7927.

Bonazzi, M., Veiga, E., Pizarro-Cerda, J., and Cossart, P. (2008). Successive post-translational modifications of E-cadherin are required for InlA-mediated internalization of Listeria monocytogenes. Cellular microbiology *10*, 2208-2222.

Borden, K.L. (2000). RING domains: master builders of molecular scaffolds? Journal of molecular biology *295*, 1103-1112.

Borden, K.L., and Freemont, P.S. (1996). The RING finger domain: a recent example of a sequence-structure family. Current opinion in structural biology *6*, 395-401.
Bordo', D., Forlani, F., Spallarossa, A., Colnaghi, R., Carpen, A., Bolognesi, M., and Pagani, S. (2001). A persulfurated cysteine promotes active site reactivity in Azotobacter vinelandii rhodanese. Biol Chem *382*, 1245-1252.
Boyer, B., Bourgeois, Y., and Poupon, M.F. (2002). Src kinase contributes to the metastatic spread of carcinoma cells. Oncogene *21*, 2347-2356.
Brinton, M.A. (2002). The molecular biology of West Nile Virus: a new invader of the western hemisphere. Annual review of microbiology *56*, 371-402.
Brown, P.H., and Schuck, P. (2006). Macromolecular size-and-shape distributions by sedimentation velocity analytical ultracentrifugation. Biophysical journal *90*, 4651-4661.
Brugge, J.S., Purchio, A.F., and Erikson, R.L. (1977). The distribution of virus-specific RNA in Rous sarcoma virus-induced hamster tumor cells. Virology *83*, 27-33.
Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S.*, et al.* (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta crystallographica Section D, Biological crystallography *54*, 905-921.
Brzovic, P.S., Rajagopal, P., Hoyt, D.W., King, M.C., and Klevit, R.E. (2001). Structure of a BRCA1-BARD1 heterodimeric RING-RING complex. Nature structural biology *8*, 833-837.
Buchwald, G., van der Stoop, P., Weichenrieder, O., Perrakis, A., van Lohuizen, M., and Sixma, T.K. (2006). Structure and E3-ligase activity of the Ring-Ring complex of polycomb proteins Bmi1 and Ring1b. Embo Journal *25*, 2465-2474.
Burke, T.R., Jr., Yao, Z.J., Liu, D.G., Voigt, J., and Gao, Y. (2001). Phosphoryltyrosyl mimetics in the design of peptide-based signal transduction inhibitors. Biopolymers *60*, 32-44.
Chayen, N.E. (1997). The role of oil in macromolecular crystallization. Structure *5*, 1269-1274.
Christofk, H.R., Vander Heiden, M.G., Wu, N., Asara, J.M., and Cantley, L.C. (2008). Pyruvate kinase M2 is a phosphotyrosine-binding protein. Nature *452*, 181-186.
Ciesla, J., Fraczyk, T., and Rode, W. (2011). Phosphorylation of basic amino acid residues in proteins: important but easily missed. Acta biochimica Polonica *58*, 137-148.
Cornilescu, G., Delaglio, F., and Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. Journal of biomolecular NMR *13*, 289-302.
Courtneidge, S.A. (2002). Role of Src in signal transduction pathways. The Jubilee Lecture. Biochemical Society transactions *30*, 11-17.
Couvineau, A., and Laburthe, M. (2012). VPAC receptors: structure, molecular pharmacology and interaction with accessory proteins. British journal of pharmacology *166*, 42-50.
Cozzone, A.J. (1988). Protein phosphorylation in prokaryotes. Annual review of microbiology *42*, 97-125.
Darnell, J.E., Jr. (1997). STATs and gene regulation. Science *277*, 1630-1635.

Davis, M.A., Ireton, R.C., and Reynolds, A.B. (2003). A core function for p120-catenin in cadherin turnover. The Journal of cell biology *163*, 525-534.
DeClue, J.E., Sadowski, I., Martin, G.S., and Pawson, T. (1987). A conserved domain regulates interactions of the v-fps protein-tyrosine kinase with the host cell. Proceedings of the National Academy of Sciences of the United States of America *84*, 9064-9068.
Del Rosario, A.M., and White, F.M. (2010). Quantifying oncogenic phosphotyrosine signaling networks through systems biology. Current opinion in genetics & development *20*, 23-30.
Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. Journal of biomolecular NMR *6*, 277-293.
Depetris, R.S., Hu, J., Gimpelevich, I., Holt, L.J., Daly, R.J., and Hubbard, S.R. (2005). Structural basis for inhibition of the insulin receptor by the adaptor protein Grb14. Molecular cell *20*, 325-333.
Deribe, Y.L., Pawson, T., and Dikic, I. (2010). Post-translational modifications in signal integration. Nat Struct Mol Biol *17*, 666-672.
Deshaies, R.J., and Joazeiro, C.A. (2009). RING domain E3 ubiquitin ligases. Annual review of biochemistry *78*, 399-434.
Dou, H., Buetow, L., Hock, A., Sibbet, G.J., Vousden, K.H., and Huang, D.T. (2012a). Structural basis for autoinhibition and phosphorylation-dependent activation of c-Cbl. Nat Struct Mol Biol *19*, 184-192.
Dou, H., Buetow, L., Sibbet, G.J., Cameron, K., and Huang, D.T. (2012b). BIRC7-E2 ubiquitin conjugate structure reveals the mechanism of ubiquitin transfer by a RING dimer. Nat Struct Mol Biol *19*, 876-883.
Doublie, S. (1997). Preparation of selenomethionyl proteins for phase determination. Methods in enzymology *276*, 523-530.
Eijkelenboom, A., and Burgering, B.M. (2013). FOXOs: signalling integrators for homeostasis maintenance. Nature reviews Molecular cell biology *14*, 83-97.
Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta crystallographica Section D, Biological crystallography *60*, 2126-2132.
Evans, J.V., Ammer, A.G., Jett, J.E., Bolcato, C.A., Breaux, J.C., Martin, K.H., Culp, M.V., Gannett, P.M., and Weed, S.A. (2012). Src binds cortactin through an SH2 domain cystine-mediated linkage. Journal of cell science *125*, 6185-6197.
Farooq, A., and Zhou, M.M. (2004). PTB or not to be: promiscuous, tolerant and Bizarro domains come of age. IUBMB life *56*, 547-557.
Fesik, S.W., Eaton, H.L., Olejniczak, E.T., Zuiderweg, E.R.P., McIntosh, L.P., and Dahlquist, F.W. (1990). 2D and 3D NMR spectroscopy employing carbon-13/carbon-13 magnetization transfer by isotropic mixing. Spin system identification in large proteins. Journal of the American Chemical Society *112*, 886-888.
Figueroa, A., Kotani, H., Toda, Y., Mazan-Mamczarz, K., Mueller, E.C., Otto, A., Disch, L., Norman, M., Ramdasi, R.M., Keshtgar, M*., et al.* (2009). Novel roles of hakai in cell proliferation and oncogenesis. Molecular biology of the cell *20*, 3533-3542.
Filippakopoulos, P., Muller, S., and Knapp, S. (2009). SH2 domains: modulators of nonreceptor tyrosine kinase activity. Current opinion in structural biology *19*, 643-649.

Flint, J., Nurizzo, D., Harding, S.E., Longman, E., Davies, G.J., Gilbert, H.J., and Bolam, D.N. (2004). Ligand-mediated dimerization of a carbohydrate-binding molecule reveals a novel mechanism for protein-carbohydrate recognition. Journal of molecular biology *337*, 417-426.
Forman-Kay, J.D., and Pawson, T. (1999). Diversity in protein recognition by PTB domains. Current opinion in structural biology *9*, 690-695.
Frame, M.C. (2004). Newest findings on the oldest oncogene; how activated src does it. Journal of cell science *117*, 989-998.
Fujita, Y., Krause, G., Scheffner, M., Zechner, D., Leddy, H.E., Behrens, J., Sommer, T., and Birchmeier, W. (2002). Hakai, a c-Cbl-like protein, ubiquitinates and induces endocytosis of the E-cadherin complex. Nature cell biology *4*, 222-231.
Gao, M., and Karin, M. (2005). Regulating the regulators: control of protein ubiquitination and ubiquitin-like modifications by extracellular stimuli. Molecular cell *19*, 581-593.
Ghosh, S., and Karin, M. (2002). Missing pieces in the NF-kappaB puzzle. Cell *109 Suppl*, S81-96.
Goddard, T.D., and Kneller, D.G. (2004). SPARKY 3. University of California, San Francisco.
Gong, E.Y., Park, E., and Lee, K. (2010). Hakai acts as a coregulator of estrogen receptor alpha in breast cancer cells. Cancer science *101*, 2019-2025.
Guarino, M. (2010). Src signaling in cancer invasion. Journal of cellular physiology *223*, 14-26.
Haglund, K., Di Fiore, P.P., and Dikic, I. (2003). Distinct monoubiquitin signals in receptor endocytosis. Trends in biochemical sciences *28*, 598-603.
Haglund, K., and Dikic, I. (2005). Ubiquitylation and cell signaling. The EMBO journal *24*, 3353-3359.
Hanahan, D., and Weinberg, R.A. (2000). The hallmarks of cancer. Cell *100*, 57-70.
Heo, H.S., Kim, J.H., Lee, Y.J., Kim, S.H., Cho, Y.S., and Kim, C.G. (2005). Microarray profiling of genes differentially expressed during erythroid differentiation of murine erythroleukemia cells. Molecules and cells *20*, 57-68.
Hershko, A., and Ciechanover, A. (1998). The ubiquitin system. Annual review of biochemistry *67*, 425-479.
Hicke, L., and Dunn, R. (2003). Regulation of membrane protein transport by ubiquitin and ubiquitin-binding proteins. Annual review of cell and developmental biology *19*, 141-172.
Hochstrasser, M. (1996). Ubiquitin-dependent protein degradation. Annual review of genetics *30*, 405-439.
Hochstrasser, M. (2009). Origin and function of ubiquitin-like proteins. Nature *458*, 422-429.
Hoege, C., Pfander, B., Moldovan, G.L., Pyrowolakis, G., and Jentsch, S. (2002). RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. Nature *419*, 135-141.
Hofmann, R.M., and Pickart, C.M. (2001). In vitro assembly and recognition of Lys-63 polyubiquitin chains. The Journal of biological chemistry *276*, 27936-27943.
Holm, L., Kaariainen, S., Rosenstrom, P., and Schenkel, A. (2008). Searching protein structure databases with DaliLite v.3. Bioinformatics *24*, 2780-2781.
Hu, J., and Hubbard, S.R. (2005). Structural characterization of a novel Cbl phosphotyrosine recognition motif in the APS family of adapter proteins. The Journal of biological chemistry *280*, 18943-18949.

Huang, A.D., Hibbert, R.G., de Jong, R.N., Das, D., Sixma, T.K., and Boelens, R. (2011). Symmetry and Asymmetry of the RING-RING Dimer of Rad18. Journal of molecular biology *410*, 424-435.
Huang, D.T., Walden, H., Duda, D., and Schulman, B.A. (2004). Ubiquitin-like protein activation. Oncogene *23*, 1958-1971.
Huang, L., Kinnucan, E., Wang, G., Beaudenon, S., Howley, P.M., Huibregtse, J.M., and Pavletich, N.P. (1999). Structure of an E6AP-UbcH7 complex: insights into ubiquitination by the E2-E3 enzyme cascade. Science *286*, 1321-1326.
Huber, K.L., Ghosh, S., and Hardy, J.A. (2012). Inhibition of caspase-9 by stabilized peptides targeting the dimerization interface. Biopolymers *98*, 451-465.
Hunter, T. (2007). The age of crosstalk: phosphorylation, ubiquitination, and beyond. Molecular cell *28*, 730-738.
Hunter, T. (2009). Tyrosine phosphorylation: thirty years and counting. Current opinion in cell biology *21*, 140-146.
Hunter, T. (2012). Why nature chose phosphate to modify proteins. Philosophical transactions of the Royal Society of London Series B, Biological sciences *367*, 2513-2516.
Ireton, R.C., Davis, M.A., van Hengel, J., Mariner, D.J., Barnes, K., Thoreson, M.A., Anastasiadis, P.Z., Matrisian, L., Bundy, L.M., Sealy, L.*, et al.* (2002). A novel role for p120 catenin in E-cadherin function. The Journal of cell biology *159*, 465-476.
Ishiyama, N., and Ikura, M. (2012). The three-dimensional structure of the cadherin-catenin complex. Sub-cellular biochemistry *60*, 39-62.
Ishiyama, N., Lee, S.H., Liu, S., Li, G.Y., Smith, M.J., Reichardt, L.F., and Ikura, M. (2010). Dynamic and static interactions between p120 catenin and E-cadherin regulate the stability of cell-cell adhesion. Cell *141*, 117-128.
Jentsch, S., and Pyrowolakis, G. (2000). Ubiquitin and its kin: how close are the family ties? Trends in cell biology *10*, 335-342.
Johnson, B.A., and Blevins, R.A. (1994). NMR View: A computer program for the visualization and analysis of NMR data. Journal of biomolecular NMR *4*, 603-614.
Joosten, R.P., Beek, T.A.H.T., Krieger, E., Hekkelman, M.L., Hooft, R.W.W., Schneider, R., Sander, C., and Vriend, G. (2011). A series of PDB related databases for everyday needs. Nucleic Acids Res *39*, D411-D419.
Kabsch, W., and Sander, C. (1983). Dictionary of Protein Secondary Structure - Pattern-Recognition of Hydrogen-Bonded and Geometrical Features. Biopolymers *22*, 2577-2637.
Kaido, M., Wada, H., Shindo, M., and Hayashi, S. (2009). Essential requirement for RING finger E3 ubiquitin ligase Hakai in early embryonic development of Drosophila. Genes to cells : devoted to molecular & cellular mechanisms *14*, 1067-1077.
Kang, X., Zhong, N., Zou, P., Zhang, S., Jin, C., and Xia, B. (2012). Foldon unfolding mediates the interconversion between M(pro)-C monomer and 3D domain-swapped dimer. Proceedings of the National Academy of Sciences of the United States of America *109*, 14900-14905.
Kasembeli, M.M., Xu, X., and Tweardy, D.J. (2009). SH2 domain binding to phosphopeptide ligands: potential for drug targeting. Frontiers in bioscience : a journal and virtual library *14*, 1010-1022.
Kato, I., Takai, T., and Kudo, A. (2002). The pre-B cell receptor signaling for apoptosis is negatively regulated by Fc gamma RIIB. Journal of immunology *168*, 629-634.

Kavanaugh, W.M., and Williams, L.T. (1994). An alternative to SH2 domains for binding tyrosine-phosphorylated proteins. Science *266*, 1862-1865.
Kim, L.C., Song, L., and Haura, E.B. (2009). Src kinases as therapeutic targets for cancer. Nature reviews Clinical oncology *6*, 587-595.
Komander, D. (2009). The emerging complexity of protein ubiquitination. Biochemical Society transactions *37*, 937-953.
Koradi, R., Billeter, M., and Wüthrich, K. (1996). MOLMOL: A program for display and analysis of macromolecular structures. Journal of Molecular Graphics *14*, 51-55.
Krishnan, M.N., Ng, A., Sukumaran, B., Gilfoy, F.D., Uchil, P.D., Sultana, H., Brass, A.L., Adametz, R., Tsui, M., Qian, F.*, et al.* (2008). RNA interference screen for human genes associated with West Nile virus infection. Nature *455*, 242-245.
Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. Journal of molecular biology *372*, 774-797.
Lampugnani, M.G., Corada, M., Andriopoulou, P., Esser, S., Risau, W., and Dejana, E. (1997). Cell confluence regulates tyrosine phosphorylation of adherens junction components in endothelial cells. Journal of cell science *110 ( Pt 17)*, 2065-2077.
Laskowski, R.A., Macarthur, M.W., Moss, D.S., and Thornton, J.M. (1993). Procheck - a Program to Check the Stereochemical Quality of Protein Structures. J Appl Crystallogr *26*, 283-291.
Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R., and Thornton, J.M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. Journal of biomolecular NMR *8*, 477-486.
Lebon, F., and Ledecq, M. (2000). Approaches to the design of effective HIV-1 protease inhibitors. Current medicinal chemistry *7*, 455-477.
Lecuit, M. (2005). Understanding how Listeria monocytogenes targets and crosses host barriers. Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases *11*, 430-436.
Lee, S., Roy, F., Galmarini, C.M., Accardi, R., Michelon, J., Viller, A., Cros, E., Dumontet, C., and Sylla, B.S. (2004). Frameshift mutation in the Dok1 gene in chronic lymphocytic leukemia. Oncogene *23*, 2287-2297.
Liang, X., Wisniewski, D., Strife, A., Shivakrupa, Clarkson, B., and Resh, M.D. (2002). Phosphatidylinositol 3-kinase and Src family kinases are required for phosphorylation and membrane recruitment of Dok-1 in c-Kit signaling. The Journal of biological chemistry *277*, 13732-13738.
Liew, C.W., Sun, H., Hunter, T., and Day, C.L. (2010). RING domain dimerization is essential for RNF4 function. The Biochemical journal *431*, 23-29.
Lim, W.A., and Pawson, T. (2010). Phosphotyrosine signaling: evolving a new cellular communication system. Cell *142*, 661-667.
Linke, K., Mace, P.D., Smith, C.A., Vaux, D.L., Silke, J., and Day, C.L. (2008). Structure of the MDM2/MDMX RING domain heterodimer reveals dimerization is required for their ubiquitylation in trans. Cell Death Differ *15*, 841-848.
Lipkowitz, S., and Weissman, A.M. (2011). RINGs of good and evil: RING finger ubiquitin ligases at the crossroads of tumour suppression and oncogenesis. Nature reviews Cancer *11*, 629-643.
Liu, B.A., Engelmann, B.W., and Nash, P.D. (2012). The language of SH2 domain interactions defines phosphotyrosine-mediated signal transduction. FEBS letters *586*, 2597-2605.

Liu, B.A., Jablonowski, K., Raina, M., Arce, M., Pawson, T., and Nash, P.D. (2006). The human and mouse complement of SH2 domain proteins-establishing the boundaries of phosphotyrosine signaling. Molecular cell *22*, 851-868.
Liu, B.A., and Nash, P.D. (2012). Evolution of SH2 domains and phosphotyrosine signalling networks. Philosophical transactions of the Royal Society of London Series B, Biological sciences *367*, 2556-2573.
Liu, Y.Q., Bai, G., Zhang, H., Su, D., Tao, D.C., Yang, Y., Ma, Y.X., and Zhang, S.Z. (2010). Human RING finger protein ZNF645 is a novel testis-specific E3 ubiquitin ligase. Asian journal of andrology *12*, 658-666.
Luo, W., Slebos, R.J., Hill, S., Li, M., Brabek, J., Amanchy, R., Chaerkady, R., Pandey, A., Ham, A.J., and Hanks, S.K. (2008). Global impact of oncogenic Src on a phosphotyrosine proteome. Journal of proteome research *7*, 3447-3460.
Lupher, M.L., Jr., Rao, N., Eck, M.J., and Band, H. (1999). The Cbl protooncoprotein: a negative regulator of immune receptor signal transduction. Immunology today *20*, 375-382.
Lupher, M.L., Jr., Songyang, Z., Shoelson, S.E., Cantley, L.C., and Band, H. (1997). The Cbl phosphotyrosine-binding domain selects a D(N/D)XpY motif and binds to the Tyr292 negative regulatory phosphorylation site of ZAP-70. The Journal of biological chemistry *272*, 33140-33144.
Ma, A., and Malynn, B.A. (2012). A20: linking a complex regulator of ubiquitylation to immunity and human disease. Nature reviews Immunology *12*, 774-785.
Mace, P.D., Linke, K., Feltham, R., Schumacher, F.R., Smith, C.A., Vaux, D.L., Silke, J., and Day, C.L. (2008). Structures of the cIAP2 RING domain reveal conformational changes associated with ubiquitin-conjugating enzyme (E2) recruitment. The Journal of biological chemistry *283*, 31633-31640.
Machida, K., and Mayer, B.J. (2005a). The SH2 domain: versatile signaling module and pharmaceutical target. Bba-Proteins Proteom *1747*, 1-25.
Machida, K., and Mayer, B.J. (2005b). The SH2 domain: versatile signaling module and pharmaceutical target. Biochimica et biophysica acta *1747*, 1-25.
Manelyte, L., Urbanke, C., Giron-Monzon, L., and Friedhoff, P. (2006). Structural and functional analysis of the MutS C-terminal tetramerization domain. Nucleic Acids Res *34*, 5270-5279.
Martin, G.S. (2001). The hunting of the Src. Nature reviews Molecular cell biology *2*, 467-475.
Mashima, R., Hishida, Y., Tezuka, T., and Yamanashi, Y. (2009). The roles of Dok family adapters in immunoreceptor signaling. Immunological reviews *232*, 273-285.
McMillan, K., Adler, M., Auld, D.S., Baldwin, J.J., Blasko, E., Browne, L.J., Chelsky, D., Davey, D., Dolle, R.E., Eagen, K.A*., et al.* (2000). Allosteric inhibitors of inducible nitric oxide synthase dimerization discovered via combinatorial chemistry. Proceedings of the National Academy of Sciences of the United States of America *97*, 1506-1511.
Meng, W., Sawasdikosol, S., Burakoff, S.J., and Eck, M.J. (1999). Structure of the amino-terminal domain of Cbl complexed to its binding site on ZAP-70 kinase. Nature *398*, 84-90.
Mengaud, J., Ohayon, H., Gounon, P., Mege, R.M., and Cossart, P. (1997). Grand entry for Listeria. Gastroenterology *112*, 1045-1046.
Metzger, M.B., Hristova, V.A., and Weissman, A.M. (2012). HECT and RING finger families of E3 ubiquitin ligases at a glance. Journal of cell science *125*, 531-537.

Mukherjee, M., Chow, S.Y., Yusoff, P., Seetharaman, J., Ng, C., Sinniah, S., Koh, X.W., Asgar, N.F., Li, D., Yim, D.*, et al.* (2012). Structure of a novel phosphotyrosine-binding domain in Hakai that targets E-cadherin. The EMBO journal *31*, 1308-1319.
Nemorin, J.G., Laporte, P., Berube, G., and Duplay, P. (2001). p62dok negatively regulates CD2 signaling in Jurkat cells. Journal of immunology *166*, 4408-4415.
Ng, C., Jackson, R.A., Buschdorf, J.P., Sun, Q., Guy, G.R., and Sivaraman, J. (2008). Structural basis for a novel intrapeptidyl H-bond and reverse binding of c-Cbl-TKB domain substrates. The EMBO journal *27*, 804-816.
Niu, Y., Roy, F., Saltel, F., Andrieu-Soler, C., Dong, W., Chantegrel, A.L., Accardi, R., Thepot, A., Foiselle, N., Tommasino, M.*, et al.* (2006). A nuclear export signal and phosphorylation regulate Dok1 subcellular localization and functions. Molecular and cellular biology *26*, 4288-4301.
Noh, S.J., Baek, H.A., Park, H.S., Jang, K.Y., Moon, W.S., Kang, M.J., Lee, D.G., Kim, M.H., Lee, J.H., and Chung, M.J. (2013). Expression of SIRT1 and cortactin is associated with progression of non-small cell lung cancer. Pathology, research and practice *209*, 365-370.
Oshida, K., Maeda, A., Kitsukawa, M., Suga, S., Iwano, S., Miyoshi, T., and Miyamoto, Y. (2011). Novel gene markers of immunosuppressive chemicals in mouse lymph node assay. Toxicology letters *205*, 79-85.
Ottiger, M., Delaglio, F., and Bax, A. (1998). Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. J Magn Reson *131*, 373-378.
Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Method Enzymol *276*, 307-326.
Park, Y.C., Burkitt, V., Villa, A.R., Tong, L., and Wu, H. (1999). Structural basis for self-association and receptor recognition of human TRAF2. Nature *398*, 533-538.
Pawson, T. (2002). Regulation and targets of receptor tyrosine kinases. European journal of cancer *38 Suppl 5*, S3-10.
Pawson, T., and Nash, P. (2003). Assembly of cell regulatory systems through protein interaction domains. Science *300*, 445-452.
Pece, S., and Gutkind, J.S. (2002). E-cadherin and Hakai: signalling, remodeling or destruction? Nature cell biology *4*, E72-74.
Pelzer, C., Kassner, I., Matentzoglu, K., Singh, R.K., Wollscheid, H.P., Scheffner, M., Schmidtke, G., and Groettrup, M. (2007). UBE1L2, a novel E1 enzyme specific for ubiquitin. The Journal of biological chemistry *282*, 23010-23014.
Penengo, L., Rubin, C., Yarden, Y., and Gaudino, G. (2003). c-Cbl is a critical modulator of the Ron tyrosine kinase receptor. Oncogene *22*, 3669-3679.
Peschard, P., Ishiyama, N., Lin, T., Lipkowitz, S., and Park, M. (2004). A conserved DpYR motif in the juxtamembrane domain of the Met receptor family forms an atypical c-Cbl/Cbl-b tyrosine kinase binding domain binding site required for suppression of oncogenic activation. The Journal of biological chemistry *279*, 29565-29571.
Petroski, M.D., and Deshaies, R.J. (2005). Function and regulation of cullin-RING ubiquitin ligases. Nature reviews Molecular cell biology *6*, 9-20.
Pickart, C.M. (2001). Mechanisms underlying ubiquitination. Annual review of biochemistry *70*, 503-533.
Polekhina, G., House, C.M., Traficante, N., Mackay, J.P., Relaix, F., Sassoon, D.A., Parker, M.W., and Bowtell, D.D. (2002). Siah ubiquitin ligase is structurally related to TRAF and modulates TNF-alpha signaling. Nature structural biology *9*, 68-75.

Ptacek, J., and Snyder, M. (2006). Charging it up: global analysis of protein phosphorylation. Trends in genetics : TIG *22*, 545-554.
Purchio, A.F., Erikson, E., Brugge, J.S., and Erikson, R.L. (1978). Identification of a polypeptide encoded by the avian sarcoma virus src gene. Proceedings of the National Academy of Sciences of the United States of America *75*, 1567-1571.
Rao, P., and Mufson, R.A. (1995). A membrane proximal domain of the human interleukin-3 receptor beta c subunit that signals DNA synthesis in NIH 3T3 cells specifically binds a complex of Src and Janus family tyrosine kinases and phosphatidylinositol 3-kinase. The Journal of biological chemistry *270*, 6886-6893.
Ren, G., Crampton, M.S., and Yap, A.S. (2009). Cortactin: Coordinating Adhesion and the Actin Cytoskeleton at Cellular Protrusions. Cell Motil Cytoskel *66*, 865-873.
Rodriguez-Rigueiro, T., Valladares-Ayerbes, M., Haz-Conde, M., Aparicio, L.A., and Figueroa, A. (2011a). Hakai reduces cell-substratum adhesion and increases epithelial cell invasion. BMC cancer *11*, 474.
Rodriguez-Rigueiro, T., Valladares-Ayerbes, M., Haz-Conde, M., Blanco, M., Aparicio, G., Fernandez-Puente, P., Blanco, F.J., Lorenzo, M.J., Aparicio, L.A., and Figueroa, A. (2011b). A novel procedure for protein extraction from formalin-fixed paraffin-embedded tissues. Proteomics *11*, 2555-2559.
Rothschild, B.L., Shim, A.H., Ammer, A.G., Kelley, L.C., Irby, K.B., Head, J.A., Chen, L., Varella-Garcia, M., Sacks, P.G., Frederick, B.*, et al.* (2006). Cortactin overexpression regulates actin-related protein 2/3 complex activity, motility, and invasion in carcinomas with chromosome 11q13 amplification. Cancer research *66*, 8017-8025.
Sadowski, I., Stone, J.C., and Pawson, T. (1986). A noncatalytic domain conserved among cytoplasmic protein-tyrosine kinases modifies the kinase function and transforming activity of Fujinami sarcoma virus P130gag-fps. Molecular and cellular biology *6*, 4396-4408.
Sakamoto, M., Takamura, M., Ino, Y., Miura, A., Genda, T., and Hirohashi, S. (2001). Involvement of c-Src in carcinoma cell motility and metastasis. Japanese journal of cancer research : Gann *92*, 941-946.
Sanchez-Vallet, A., Saleem-Batcha, R., Kombrink, A., Hansen, G., Valkenburg, D.J., Thomma, B.P., and Mesters, J.R. (2013). Fungal effector Ecp6 outcompetes host immune receptor for chitin binding through intrachain LysM dimerization. eLife *2*, e00790.
Sawyer, T.K., Bohacek, R.S., Dalgarno, D.C., Eyermann, C.J., Kawahata, N., Metcalf, C.A., 3rd, Shakespeare, W.C., Sundaramoorthi, R., Wang, Y., and Yang, M.G. (2002). SRC homology-2 inhibitors: peptidomimetic and nonpeptide. Mini reviews in medicinal chemistry *2*, 475-488.
Scheffner, M., and Staub, O. (2007). HECT E3s and human disease. BMC biochemistry *8 Suppl 1*, S6.
Schlessinger, J., and Lemmon, M.A. (2003). SH2 and PTB domains in tyrosine kinase signaling. Science's STKE : signal transduction knowledge environment *2003*, RE12.
Schmidt, M.H., and Dikic, I. (2005). The Cbl interactome and its functions. Nature reviews Molecular cell biology *6*, 907-918.
Schulman, B.A. (2011). Twists and turns in ubiquitin-like protein conjugation cascades. Protein science : a publication of the Protein Society *20*, 1941-1954.
Schulman, B.A., and Harper, J.W. (2009). Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways. Nature reviews Molecular cell biology *10*, 319-331.

Schwieters, C.D., Kuszewski, J.J., Tjandra, N., and Clore, G.M. (2003). The Xplor-NIH NMR molecular structure determination package. J Magn Reson *160*, 65-73.
Selbach, M., and Backert, S. (2005). Cortactin: an Achilles' heel of the actin cytoskeleton targeted by pathogens. Trends in microbiology *13*, 181-189.
Shen, Y., Naujokas, M., Park, M., and Ireton, K. (2000). InIB-dependent internalization of Listeria is mediated by the Met receptor tyrosine kinase. Cell *103*, 501-510.
Siouda, M., Yue, J., Shukla, R., Guillermier, S., Herceg, Z., Creveaux, M., Accardi, R., Tommasino, M., and Sylla, B.S. (2012). Transcriptional regulation of the human tumor suppressor DOK1 by E2F1. Molecular and cellular biology *32*, 4877-4890.
Smith, M.J., Hardy, W.R., Murphy, J.M., Jones, N., and Pawson, T. (2006). Screening for PTB domain binding partners and ligand specificity using proteome-derived NPXY peptide arrays. Molecular and cellular biology *26*, 8461-8474.
Soler-Lopez, M., Petosa, C., Fukuzawa, M., Ravelli, R., Williams, J.G., and Muller, C.W. (2004). Structure of an activated Dictyostelium STAT in its DNA-unbound form. Molecular cell *13*, 791-804.
Songyang, Z., Shoelson, S.E., Chaudhuri, M., Gish, G., Pawson, T., Haser, W.G., King, F., Roberts, T., Ratnofsky, S., Lechleider, R.J*., et al.* (1993). SH2 domains recognize specific phosphopeptide sequences. Cell *72*, 767-778.
Stein, E.G., Ghirlando, R., and Hubbard, S.R. (2003). Structural basis for dimerization of the Grb10 Src homology 2 domain. Implications for ligand specificity. The Journal of biological chemistry *278*, 13257-13264.
Sun, Q., Jackson, R.A., Ng, C., Guy, G.R., and Sivaraman, J. (2010). Additional serine/threonine phosphorylation reduces binding affinity but preserves interface topography of substrate proteins to the c-Cbl TKB domain. PloS one *5*, e12819.
Sun, Q., Ng, C., Guy, G.R., and Sivaraman, J. (2011). An adjacent arginine, and the phosphorylated tyrosine in the c-Met receptor target sequence, dictates the orientation of c-Cbl binding. FEBS letters *585*, 281-285.
Tamagnone, L., Artigiani, S., Chen, H., He, Z., Ming, G.I., Song, H., Chedotal, A., Winberg, M.L., Goodman, C.S., Poo, M*., et al.* (1999). Plexins are a large family of receptors for transmembrane, secreted, and GPI-anchored semaphorins in vertebrates. Cell *99*, 71-80.
Tamir, I., Stolpa, J.C., Helgason, C.D., Nakamura, K., Bruhns, P., Daeron, M., and Cambier, J.C. (2000). The RasGAP-binding protein p62dok is a mediator of inhibitory FcgammaRIIB signals in B cells. Immunity *12*, 347-358.
Taylor, J.D., Ababou, A., Fawaz, R.R., Hobbs, C.J., Williams, M.A., and Ladbury, J.E. (2008). Structure, dynamics, and binding thermodynamics of the v-Src SH2 domain: Implications for drug design. Proteins *73*, 929-940.
Terwilliger, T.C. (2003). Automated main-chain model building by template matching and iterative fragment extension. Acta Crystallogr D *59*, 38-44.
Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Crystallogr D *55*, 849-861.
Thiery, J.P., and Sleeman, J.P. (2006). Complex networks orchestrate epithelial-mesenchymal transitions. Nature reviews Molecular cell biology *7*, 131-142.
Thomas, S.M., and Brugge, J.S. (1997). Cellular functions regulated by Src family kinases. Annual review of cell and developmental biology *13*, 513-609.
Thoreson, M.A., Anastasiadis, P.Z., Daniel, J.M., Ireton, R.C., Wheelock, M.J., Johnson, K.R., Hummingbird, D.K., and Reynolds, A.B. (2000). Selective uncoupling

of p120(ctn) from E-cadherin disrupts strong adhesion. The Journal of cell biology *148*, 189-202.
Wenta, N., Strauss, H., Meyer, S., and Vinkemeier, U. (2008). Tyrosine phosphorylation regulates the partitioning of STAT1 between different dimer conformations. Proceedings of the National Academy of Sciences of the United States of America *105*, 9238-9243.
Wenzel, D.M., and Klevit, R.E. (2012). Following Ariadne's thread: a new perspective on RBR ubiquitin ligases. BMC biology *10*, 24.
Wilkinson, K.D. (2000). Ubiquitination and deubiquitination: targeting of proteins for degradation by the proteasome. Seminars in cell & developmental biology *11*, 141-148.
Woodring, P.J., Meisenhelder, J., Johnson, S.A., Zhou, G.L., Field, J., Shah, K., Bladt, F., Pawson, T., Niki, M., Pandolfi, P.P.*, et al.* (2004). c-Abl phosphorylates Dok1 to promote filopodia during cell spreading. The Journal of cell biology *165*, 493-503.
Wu, H., Reynolds, A.B., Kanner, S.B., Vines, R.R., and Parsons, J.T. (1991). Identification and characterization of a novel cytoskeleton-associated pp60src substrate. Molecular and cellular biology *11*, 5113-5124.
Xiao, K., Allison, D.F., Buckley, K.M., Kottke, M.D., Vincent, P.A., Faundez, V., and Kowalczyk, A.P. (2003). Cellular levels of p120 catenin function as a set point for cadherin expression levels in microvascular endothelial cells. The Journal of cell biology *163*, 535-545.
Xu, Z., Kohli, E., Devlin, K.I., Bold, M., Nix, J.C., and Misra, S. (2008). Interactions between the quality control ubiquitin ligase CHIP and ubiquitin conjugating enzymes. Bmc Struct Biol *8*.
Yaffe, M.B. (2002). Phosphotyrosine-binding domains in signal transduction. Nature reviews Molecular cell biology *3*, 177-186.
Yamada, S., Pokutta, S., Drees, F., Weis, W.I., and Nelson, W.J. (2005). Deconstructing the cadherin-catenin-actin complex. Cell *123*, 889-901.
Yamakawa, N., Tsuchida, K., and Sugino, H. (2002). The rasGAP-binding protein, Dok-1, mediates activin signaling via serine/threonine kinase receptors. The EMBO journal *21*, 1684-1694.
Ye, Y., and Rape, M. (2009). Building ubiquitin chains: E2 enzymes at work. Nature reviews Molecular cell biology *10*, 755-764.
Yeatman, T.J. (2004). A renaissance for SRC. Nature reviews Cancer *4*, 470-480.
Yeh, E.T., Gong, L., and Kamitani, T. (2000). Ubiquitin-like proteins: new wines in new bottles. Gene *248*, 1-14.
Yin, Q., Lin, S.C., Lamothe, B., Lu, M., Lo, Y.C., Hura, G., Zheng, L.X., Rich, R.L., Campos, A.D., Myszka, D.G.*, et al.* (2009). E2 interaction and dimerization in the crystal structure of TRAF6. Nat Struct Mol Biol *16*, 658-U697.
Yusoff, P., Lao, D.H., Ong, S.H., Wong, E.S.M., Lim, J., Lo, T.L., Leong, H.F., Fong, C.W., and Guy, G.R. (2002). Sprouty2 inhibits the Ras/MAP kinase pathway by inhibiting the activation of raf. Journal of Biological Chemistry *277*, 3195-3201.
Zhang, H.M., Guo, T.N., Li, X., Datta, A., Park, J.E., Yang, J., Lim, S.K., Tam, J.P., and Sze, S.K. (2010). Simultaneous Characterization of Glyco- and Phosphoproteomes of Mouse Brain Membrane Proteome with Electrostatic Repulsion Hydrophilic Interaction Chromatography. Mol Cell Proteomics *9*, 635-647.

Zhang, L., Fairall, L., Goult, B.T., Calkin, A.C., Hong, C., Millard, C.J., Tontonoz, P., and Schwabe, J.W.R. (2011). The IDOL-UBE2D complex mediates sterol-dependent degradation of the LDL receptor. Gene Dev *25*, 1262-1274.
Zhao, M., Schmitz, A.A., Qin, Y., Di Cristofano, A., Pandolfi, P.P., and Van Aelst, L. (2001). Phosphoinositide 3-kinase-dependent membrane recruitment of p62(dok) is essential for its negative effect on mitogen-activated protein (MAP) kinase activation. The Journal of experimental medicine *194*, 265-274.

**The EMBO Journal (2012) 31,** 1308–1319 | © 2012 European Molecular Biology Organization | All Rights Reserved 0261-4189/12
www.embojournal.org

THE EMBO JOURNAL

# Structure of a novel phosphotyrosine-binding domain in Hakai that targets E-cadherin

**Manjeet Mukherjee[1,6], Soah Yee Chow[2,6], Permeen Yusoff[2], J Seetharaman[3], Cherlyn Ng[2], Saravanan Sinniah[2], Xiao Woon Koh[2], Nur Farehan M Asgar[2], Dan Li[2], Daniel Yim[2], Rebecca A Jackson[2], Jingxi Yew[2], Jingru Qian[4], Audrey Iyu[2], Yoon Pin Lim[5], Xingding Zhou[1], Siu Kwan Sze[4], Graeme R Guy[2,*] and J Sivaraman[1,*]**

[1]Department of Biological Sciences, National University of Singapore, Singapore; [2]Signal Transduction Laboratory, Institute of Molecular and Cell Biology, Proteos, Singapore; [3]X4 Beamline, Brookhaven National Laboratory, Upton, NY, USA; [4]School of Biological Sciences, Nanyang Technological University, Singapore and [5]Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore

**Phosphotyrosine-binding domains, typified by the SH2 (Src homology 2) and PTB domains, are critical upstream components of signal transduction pathways. The E3 ubiquitin ligase Hakai targets tyrosine-phosphorylated E-cadherin via an uncharacterized domain. In this study, the crystal structure of Hakai (amino acids 106–206) revealed that it forms an atypical, zinc-coordinated homodimer by utilizing residues from the phosphotyrosine-binding domain of two Hakai monomers. Hakai dimerization allows the formation of a phosphotyrosine-binding pocket that recognizes specific phosphorylated tyrosines and flanking acidic amino acids of Src substrates, such as E-cadherin, cortactin and DOK1. NMR and mutational analysis identified the Hakai residues required for target binding within the binding pocket, now named the HYB domain. ZNF645 also possesses a HYB domain but demonstrates different target specificities. The HYB domain is structurally different from other phosphotyrosine-binding domains and is a potential drug target due to its novel structural features.**

*The EMBO Journal* (2012) **31,** 1308–1319. doi:10.1038/emboj.2011.496; Published online 17 January 2012
*Subject Categories*: cell & tissue architecture; signal transduction
*Keywords*: Hakai; phosphotyrosine; Src substrates; ubiquitin ligases; zinc fingers

*Corresponding author. GR Guy, Signal Transduction Laboratory, Institute of Molecular and Cell Biology, 61 Biopolis Drive, Proteos, Singapore 138673, Singapore. Tel.: +65 65869614; Fax: +65 67791117; E-mail: mcbgg@imcb.a-star.edu.sg or J Sivaraman, Department of Biological Sciences, 14 Science Drive 4, National University of Singapore, Singapore 117543, Singapore. Tel.: +65 65161163; Fax: +65 67795671; E-mail: dbsjayar@nus.edu.sg
[6]Joint first authors

**Received: 23 August 2011; accepted: 21 December 2011; published online: 17 January 2012**

## Introduction

In eukaryotic cells, phosphorylation events regulate cell signalling by providing docking sites for protein domains, such as the Src homology 2 (SH2) and phosphotyrosine-binding (PTB) domains (Forman-Kay and Pawson, 1999; Yaffe, 2002; Pawson and Nash, 2003). The SH2 was the first signalling domain to be identified and has been extensively characterized (Songyang *et al*, 1993; Forman-Kay and Pawson, 1999; Yaffe, 2002; Pawson and Nash, 2003; Liu *et al*, 2006; Filippakopoulos *et al*, 2009). The SH2 is a dedicated phosphotyrosine-binding domain and plays a critical role in signal transduction, hence making it a target for drug development (Pawson and Nash, 2003; Machida and Mayer, 2005; Taylor *et al*, 2008; Kasembeli *et al*, 2009). Binding specificity of SH2 domains is generally conferred by the sequences flanking the C-terminus of the phosphotyrosine (pTyr), and motif recognition is usually relatively inflexible. The other major class of pTyr-binding domain is the PTB domain. The specificity of binding to the PTB domain is conferred typically by residues on the target that are N-terminal to the pTyr. However, the PTB domain also recognizes non-pTyr motifs (Forman-Kay and Pawson, 1999; Yaffe, 2002; Pawson and Nash, 2003; Farooq and Zhou, 2004; Smith *et al*, 2006). Atypical phosphotyrosine-binding domains have also been detected in PKCδ and the human M2 pyruvate kinase (PKM2) (Benes *et al*, 2005; Christofk *et al*, 2008).

In 2002, Fujita *et al* (2002) discovered a novel ubiquitin E3 ligase protein that targeted pTyr sites on E-cadherin. The E3 ligase, Hakai, possesses three domains: a RING domain, a short pTyr recognition sequence and a proline-rich domain (Fujita *et al*, 2002). Hakai is involved in the regulation of cell adhesion, cell migration and embryogenesis (Figueroa *et al*, 2009; Kaido *et al*, 2009; Gong *et al*, 2010). Among the reported protein interactions of Hakai, its association with and ubiquitination of E-cadherin upon Src activation is the best characterized (Fujita *et al*, 2002).

Based on molecular modelling, Fujita *et al* (2002) assumed the pTyr-binding domain of Hakai to be a derivative SH2 domain. Our previous experience analysing the SH2 domain of a similar E3 ubiquitin ligase, c-Cbl (Ng *et al*, 2008) led us to examine the nature of the Hakai pTyr-binding domain to provide structural insights into its interaction with E-cadherin. In this study, we report that the Hakai pTyr-binding (HYB) domain consists of a homodimer formed at a structurally novel interface. Each monomer consists of two zinc-finger domains: a RING domain and a minimum pTyr-binding domain that incorporates a novel, atypical zinc coordination motif. Both domains play key roles in dimerization. The HYB domain is therefore composed of four zinc-binding domains cooperating to bind pTyr residues surrounded by acidic amino acids. Whereas the RING domain appears in other proteins, the atypical zinc-binding domain component is a novel protein fold that incorporates an intertwined config-